«Solo Sean Carroll podría introducir la relatividad general de Einstein de una manera tan luminosa y directa». **Carlo Rovelli**

Las ideas fundamentales del Universo

Nº1 best seller *The New York Times*

Espacio, tiempo y movimiento

Sean Carroll

Profesor de Física en el Instituto de Tecnología de California (Caltech)

arpa

«Solo Sean Carroll podría introducir la relatividad general de Einstein de una manera tan luminosa y directa». **Carlo Rovelli**

Las ideas fundamentales del Universo

Nº1 best seller *The New York Times*

Espacio, tiempo y movimiento

Sean Carroll

Profesor de Física en el Instituto de Tecnología de California (Caltech)

arpa

LAS IDEAS FUNDAMENTALES

DEL UNIVERSO

Título original: The Biggest Ideas in the Universe:

Space, Time, and Motion

© del texto: Sean Carroll, 2022

© de la traducción: Jordi Giménez Samanes, 2023

© de esta edición: Arpa & Alfil Editores, S. L.

Primera edición: marzo de 2023

ISBN: 978-84-19558-03-9

Depósito legal: B 4409-2023

Diseño de cubierta: Anna Juvé

Maquetación: Nèlia Creixell, Judit Pulido

Producción del ePub: booqlab

Arpa

Manila, 65

08034 Barcelona

arpaeditores.com

Reservados todos los derechos.

Ninguna parte de esta publicación puede ser reproducida, almacenada o transmitida por ningún medio sin permiso del editor.

Sean Carroll

LAS IDEAS FUNDAMENTALES DEL UNIVERSO

ESPACIO, TIEMPO Y MOVIMIENTO

Traducción de Jordi Giménez Samanes

arpa

A Jennifer

ÍNDICE

Introducción

UNO Conservación

DOS Cambio

TRES Dinámica

CUATRO Espacio

CINCO Tiempo

SEIS Espacio-tiempo

SIETE Geometría

OCHO Gravedad

NUEVE Agujeros negros

Apéndices

Apéndice A: Funciones, derivadas e integrales

Apéndice B: Conexiones y curvatura

INTRODUCCIÓN

Mi sueño sería vivir en un mundo en el que la mayoría de las personas estuvieran bien informadas y albergaran opiniones apasionadas en torno a la física moderna. En el que, a la salida de un día duro en el trabajo, nos reuniéramos en el pub con unos amigos y nos pusiéramos a discutir sobre cuál es nuestro candidato preferido a materia oscura, o sobre interpretaciones contrapuestas en torno a la mecánica cuántica. Un mundo en el que, mientras los niños corretearan en una fiesta de cumpleaños, un padre dijera: «No entiendo cómo hay personas que piensan que tiene que haber nuevas partículas próximas a la escala electrodébil»; y otro replicara de inmediato: «Entonces ¿cómo quieres abordar el problema de jerarquía?». Al fin y al cabo, la gente tiene una opinión acerca de la economía de oferta o la teoría crítica de la raza. ¿Por qué no acerca de la cosmología inflacionaria y la teoría de supercuerdas?

Este no es en absoluto el mundo en que vivimos. En mayor medida aún que la mayor parte de las demás disciplinas académicas, la física es un terreno cultivado por y para especialistas. Sus profesionales hablan entre sí en una jerga altamente especializada, en la cual prevalecen conceptos matemáticos de los que la mayoría de la gente jamás ha oído hablar, mucho menos domina. Hay razones lógicas para que esto sea así, pero no tiene por qué serlo obligatoriamente. La situación se debe en gran medida a la forma en que los físicos tienden a compartir su conocimiento con el resto del mundo.

Si es usted una persona no experta interesada en aprender cosas acerca de la física moderna, tiene básicamente dos opciones. Una es no salirse de una explicación de nivel popular, por la que poder aprender algo acerca de los conceptos importantes, pero sin ahondar en los detalles técnicos o matemáticos. Puede leer libros, asistir a conferencias, ver vídeos, escuchar pódcast. Lo bueno es que contamos ciertamente con un boyante ecosistema de tales recursos, por lo que es posible aprender bastante, aunque de un modo poco sistemático. En el fondo, no obstante, uno sabe que así no alcanzará el fundamento. Lo que obtendrá serán imágenes y metáforas,

burdas traducciones al lenguaje corriente de la esencia matemática subyacente. Puede llegar a recorrer una distancia impresionante por este camino, pero siempre se perderá algo vital.

El otro camino es hacerse estudiante de física: bien yendo literalmente a la universidad, bien reuniendo los libros de texto y los recursos online idóneos. En el proceso, necesitará conseguir buenas competencias matemáticas: en cálculo y ecuaciones diferenciales por encima de todo, pero en aspectos como análisis vectorial, números complejos, álgebra lineal y otros más. El viaje será gratificante, pero frustrantemente lento. En general, se requiere cursar al menos un año de materias introductorias antes de que el estudiante oiga hablar siquiera de relatividad o de mecánica cuántica. Y la mayoría de estudiantes de física pueden obtener una licenciatura —o incluso realizar todos los cursos conducentes al título de doctor— sin aprender nada de física de las partículas, agujeros negros o cosmología. Tales exquisiteces están reservadas en exclusiva a especialistas de ramas más específicas.

La distancia que separa aprender física con un interés de aficionado, confiando en metáforas y oscuras traducciones, de convertirse en un experto acreditado, moviéndose con comodidad entre ecuaciones de complejidad intimidatoria, es amplia, pero no insalvable. Solo porque yo no quiera ser un piloto de carreras profesional, no significa que vayan a prohibirme conducir. Sin duda existe un camino por el que adentrarse en algo de la esencia auténtica de la física moderna —aunque ello comporte contemplar unas pocas ecuaciones— sin tener que afanarse trabajosamente a lo largo de años de estudios académicos programados.

¡Ha venido al lugar adecuado!

Las ideas fundamentales del universo se centra en la idea de que es posible aprender física moderna de verdad, ecuaciones incluidas y todo lo demás, aun si usted es más aficionado que profesional y tiene toda la intención de seguir siendo así. Está pensado para personas que no tienen más experiencia en matemáticas que la del álgebra del instituto, pero que están dispuestas a detenerse en una ecuación y pensar acerca de lo que significa. Si está dispuesto a realizar este pequeño esfuerzo de pensamiento, un nuevo mundo se abrirá ante sus ojos.

***

Pasa una cosa con las ecuaciones: que no tienen por qué asustarnos tanto. Constituyen simplemente un modo de resumir de forma compacta una relación entre diferentes cantidades. Una cosa es que nos digan, de acuerdo con la teoría de la relatividad general de Einstein: «la masa y la energía causan una curvatura en el espacio-tiempo»; y otra muy distinta es que nos den la ecuación de Einstein:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}$$

La frase en nuestra lengua corriente nos infunde algo así como un sentimiento acerca de lo que trata la relatividad general, pero la ecuación nos dice lo que realmente es, en términos precisos y sin ambigüedad. Podemos leer todas las palabras que queramos, pero hasta que no entendamos esta ecuación, no entenderemos de verdad la teoría de Einstein.

El problema está en que esta ecuación nos resulta totalmente opaca si no sabemos qué significan esos símbolos. Es un galimatías. Para que nuestra mente la capte, necesitamos entender el papel individual que desempeña cada uno de los números y letras, incluidos los subíndices μ y υ, letras procedentes del alfabeto griego, por si fuera poco. Hay buenas razones por las que un estudiante medio de física tarda años en recorrer tan largo trecho.

Pero usted lo hará leyendo este libro. Cuando llegue al capítulo 8, entenderá lo que significan todos los símbolos de la ecuación de Einstein, cómo encajan unos con otros y qué es lo que nos dicen acerca del espacio-tiempo y la gravedad. Por mucho que la ecuación incluya letras griegas, entenderla es enormemente más fácil que, pongamos, aprender a hablar y escribir griego moderno.

La mayor parte de los libros de divulgación dan por sentado que usted no quiere hacer el esfuerzo de seguir las ecuaciones. Los libros de texto, por otro lado, asumen que usted no solo quiere entender las ecuaciones, sino resolverlas. Y resolver estas ecuaciones requiere un trabajo, una práctica y un aprendizaje enormemente mayores que «meramente» entenderlas.

Permítame insistir en esta distinción entre resolver y entender, porque será la clave para el progreso, tan notablemente rápido, que seremos capaces de realizar. La ecuación de Einstein no solo relaciona un conjunto específico de masa y energía con la curvatura de un espacio-tiempo concreto. Se trata de una relación perfectamente general, del tipo: «usted deme cierta distribución de masa y de energía, y yo le diré cómo se curva el espacio-tiempo debido a ella». Cumplir tal promesa es a lo que nos referimos con «resolver la ecuación».

A veces resolver una ecuación es fácil: si la ecuación es $x=y^2$, y nos dicen que $y=2$, la solución es $x=4$. No es tan difícil. Pero las ecuaciones de física del mundo real son algo más complicadas, e incluyen nociones de cálculo (las matemáticas del cambio continuo) y otros conceptos avanzados. Resolver tales ecuaciones puede convertirse en una ocupación a tiempo completo para los profesionales de la física. Es muy sensato, por tanto, que en su educación ocupe una parte muy amplia aprender a resolver ecuaciones. Cualquier estudiante de física le dirá que la parte más difícil de los años que lleva en la facultad no es seguir las conferencias a las que asiste, sino resolver las tandas de problemas que los profesores no paran de repartir, como si los estudiantes no tuvieran nada más que hacer los fines de semana.

Aquí, en Las ideas fundamentales del universo, no vamos a enseñarle cómo resolver las ecuaciones, pero sí que aprenderá a entender las ecuaciones, incluso las que se consideran relativamente avanzadas para el nivel de los manuales de física. Porque resulta que es enormemente más fácil. Tales libros se sustentan en la creencia de que las ideas de los físicos modernos —las de verdad, no las versiones metafóricas rebajadas— pueden ser accesibles a quienquiera que esté dispuesto a realizar un poco de esfuerzo para pensar acerca de las ecuaciones y lo que significan.

***

Muy bien, pero ¿cuáles son esas ideas de las que estamos hablando? Hay muchas, como puede imaginar. Las suficientes como para que hayamos repartido el material en una serie en tres partes: Espacio, tiempo y movimiento; Cuantos y campos; y Complejidad y emergencia. Este formato a modo de trilogía se ha demostrado exitoso para El señor de los anillos y otras series populares.

El libro que sostiene en sus manos, Espacio, tiempo y movimiento, está centrado en el marco conceptual de la física clásica abanderada por Isaac Newton, que fue el dominante hasta la revolución cuántica del siglo XX. Pero no tema, no vamos a gastar mucho tiempo hablando de poleas y planos inclinados, por importantes que sean. El ámbito de la física clásica incluye profundas cuestiones acerca de la naturaleza del espacio, el tiempo y el cambio, por lo que no tendrá que asustarnos salpicar de algunas consideraciones filosóficas nuestras ecuaciones. Incluye asimismo la teoría de la relatividad, hasta alcanzar las ideas de Einstein acerca del espacio-tiempo curvo, junto con consecuencias tales como los agujeros negros. De modo que este libro con ideas que tienen siglos de antigüedad, pero que nos conducirán hasta conceptos de los que se ocupa la investigación moderna.

En Cuantos y campos trataremos de ideas cuánticas tan excitantes como el entrelazamiento y el gato de Schrödinger, pero sobre todo tendremos oportunidad de aprender cosas acerca de la teoría cuántica de campos y la física de partículas, el mejor planteamiento moderno sobre las leyes fundamentales de la naturaleza. En el episodio final, Complejidad y emergencia, es donde admitimos que el mundo no está hecho tan solo de dos o tres partículas. Suceden cosas interesantes cuando los sistemas consisten en un gran número de partes en movimiento.

Son muchos conceptos, pero casi todos ellos entran en el reino de la física y ámbitos afines. No se trata de desdeñar las ideas, igualmente grandes e importantes, de otros ámbitos de la ciencia (o del arte y las humanidades, dicho sea de paso), pero en algún lugar tenemos que trazar la línea.

Otra línea que hemos tenido que trazar separa las «ideas sobre las que tenemos buenas razones para creer que son ciertas» de las «especulaciones prometedoras». Mientras que los manuales de física tienden a no apartarse de las ideas cuya utilidad está establecida, las perspectivas más populares se entregan alegremente a conceptos que son todavía enteramente hipotéticos. El procedimiento es perfectamente sensato, los investigadores se pasan la mayor parte del tiempo en la frontera, pensando en posibilidades que aún no han pasado a formar parte de la tradición. Nuestro objetivo es el de asumir las ideas sobre las que tenemos excelentes razones para pensar que seguirán formando parte de las herramientas de trabajo de los profesionales de la física dentro de cien años.

***

Es muy grato para mí poder expresar mi agradecimiento a la enorme ayuda recibida a lo largo del camino. Scott Aaronson, Justin Clarke-Doane y Matt Strassler me han proporcionado un intercambio inestimable y me han salvado de algo más que unas pocas formas de expresión desacertadas. Jason Torchinsky es el autor de las bellas ilustraciones. Mi editor, Stephen Morrow, me ha prestado todo su apoyo y toda su perspicacia, como siempre, y mi agente, Katinka Matson, me ha ayudado a dar forma a un proyecto que era complicado. El concurso de Alice Dalrymple, Tiffany Estreicher, Dora Mak, Nakeesha Warner y Melanie Muto ha sido vital para el proceso de producción. La idea del libro surgió de una serie de vídeos que realicé durante la pandemia de la COVID-19, inspirado por las clases online de escritura dramática impartidas por mi amiga Lauren Gunderson. Y, por supuesto, no podré agradecer lo bastante a Jennifer Ouellette su asesoramiento en la redacción, su apoyo moral, y tantas cosas más.

Si desea ver los vídeos, o consultar otros materiales complementarios, puede visitar: presposterousuniverse.com/biggestideas/.

UNO

CONSERVACIÓN

Mire a su alrededor. Si es como la mayoría de las personas, tiene usted un cuerpo. Está ubicado en algún lugar. Hay indudables posibilidades de que esté rodeado por una variedad de otros objetos, ubicados en otros tantos lugares. Mesas, sillas, un suelo, un techo, paredes, árboles quizás o cierta extensión de agua, si está en el exterior. Todos esos objetos existen, con sus ubicaciones y propiedades, y esas ubicaciones y propiedades pueden cambiar con el tiempo. Puede acercar la silla a la pared, o desplazarla más lejos. Puede beber un vaso de agua, y su cuerpo asimilará la substancia. Y si, en lugar de eso, coloca el vaso sobre una mesa y lo deja ahí abandonado, el agua terminará por evaporarse en el aire.

Así es como pensamos acerca del mundo desde una perspectiva inmediata, a escala humana. Hay materia, que está ubicada en el espacio. (Por «espacio» no entendemos «el espacio exterior», sino simplemente el reino tridimensional por el que se mueven las cosas). Esa materia puede cambiar, o puede permanecer constante a lo largo del tiempo. La física es el estudio de todo ese mundo material, así como de su comportamiento, al nivel más básico que podamos pensar. ¿Qué es toda esa materia, en realidad? ¿Cómo se relacionan los diferentes objetos entre sí? ¿Cómo cambian con el paso del tiempo? ¿Qué es el «tiempo»? ¿Y qué es el «espacio», si lo analizamos?

Una de las características que más nos hacen disfrutar de la física es lo rápidamente que pasamos de las observaciones mundanas —¡mira cómo reaccionan esos objetos!— a las cuestiones más profundas acerca de la naturaleza de la realidad. La clave está en que las cosas no suceden sin más: todos los hechos siguen ciertos patrones. Tales patrones es a lo que llamamos las leyes de la física, y nuestra labor es desvelarlos.

El más simple patrón de todos es el hecho de que ciertas cosas permanecen constantes aunque pase el tiempo. Contemplar este rasgo básico de la realidad es ya un gran punto de partida para nuestras investigaciones, que no tardarán mucho en desbocarse.

PREDICTIBILIDAD

Damos por sentado que el mundo que nos rodea es mínimamente predecible. Si hay una mesa en la habitación y le damos la espalda un segundo, esperamos que la mesa siga ahí cuando nos volvamos. Si dejamos una manzana encima de la mesa, esperamos que la mesa la sostenga, en lugar de que la manzana atraviese la mesa y caiga al suelo. Por mucho que nos quejemos de lo difícil que es predecir el tiempo o el resultado de las próximas elecciones, nos sentiríamos impresionados al comprobar el grado de fiabilidad de la predictibilidad.

La física es posible gracias a esta predictibilidad. Puede que no sea absoluta, pero podemos hasta cierto punto anticipar lo que va a suceder a continuación en el mundo sabiendo lo que está sucediendo ahora mismo. El tipo más básico de predictibilidad es la conservación, el hecho de que hay cosas que no cambian.

Conservación es la forma en que los físicos se refieren a «permanecer constante a lo largo del tiempo». Habrá oído decir que la energía se conserva, por ejemplo. La energía no es un tipo de sustancia, como el agua o la tierra. Es una propiedad que tienen las cosas, dependiendo de lo que son y de la situación en que están. No existe un «flujo de energía» que fluye de un lugar a otro. Simplemente hay objetos con sus posiciones, sus velocidades y otras propiedades, y a esos objetos podemos asociarles cierta cantidad de energía en relación con tales hechos.

Un objeto puede tener energía porque se esté moviendo, porque esté ubicado en una posición elevada, porque esté caliente, por su masa, por su carga eléctrica, o por otras razones. Bajo las circunstancias pertinentes, tales

formas de energía pueden convertirse y reconvertirse unas en otras. La energía que tiene una copa de vino por el mero hecho de estar encima de una mesa, puede, si alguien empuja la copa y la tira por el borde, convertirse rápidamente en energía cinética mientras cae, y luego en calor, en ruido y en otras formas de energía disipada al romperse contra el suelo. La conservación de la energía recoge simplemente la idea de que la energía total, resultante de la suma de todas sus formas individuales, permanece constante a lo largo de todo el proceso.

(Un momento, ¿estamos razonando en círculo? ¿No estaremos meramente inventando un conjunto de cantidades que suman un número constante por definición, a lo que hemos llamado «energía», y felicitándonos por haber descubierto una ley de la física? No. Hay una forma independiente de definir la energía, y luego mostrar que se conserva, a partir del hecho de que las leyes de la física no cambian con el tiempo. Pero ha planteado la clase de pregunta adecuada).

Tan simple idea como podríamos imaginar, que hay una cantidad que no cambia, dice lo mismo por tiempo que pase. Pero la conservación de la energía y otras cantidades no es tan solo una posición amable y tranquilizadora desde la que abordar todo el estudio de la física. Desde un punto de vista lógico, es el lugar adecuado, ya que una comprensión cabal de la conservación representó el primer paso para la transición de la ciencia premoderna a la moderna.

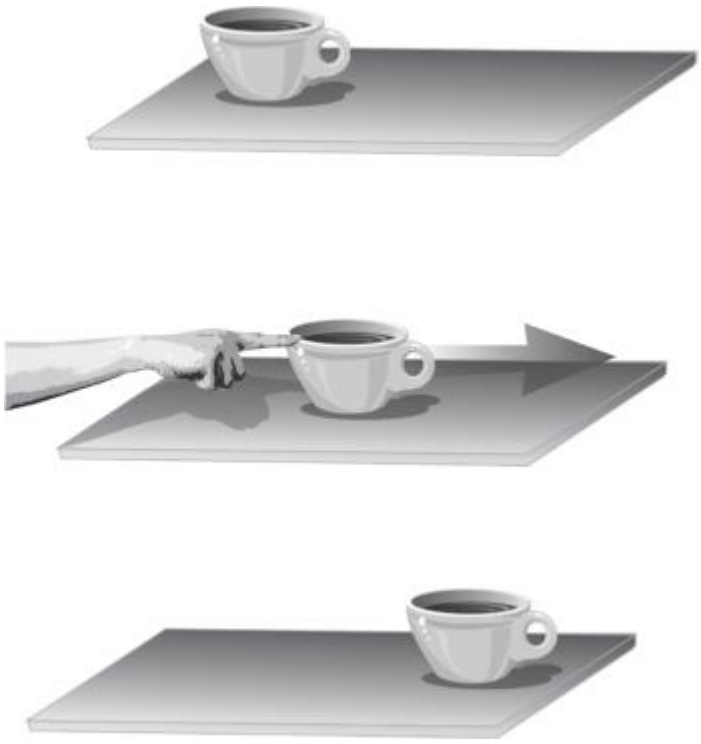
DE NATURALEZAS A PATRONES

Póngase en la mentalidad de los seres humanos que trataban de comprender el mundo antes de la llegada de la física moderna. El filósofo griego Aristóteles suele ser el elegido como ejemplo, si bien otros pensadores antiguos debieron razonar de un modo similar. Para simplificar notablemente un complejo y sutil conglomerado de ideas, Aristóteles dividió el modo en que las cosas se mueven entre movimiento «natural» e «innatural» (o «violento»). Concebía el mundo fundamentalmente en

términos teleológicos, orientado hacia un fin. Los objetos tienen un lugar natural o una condición natural en que estar, y tienden a moverse hacia ese lugar. Una piedra caerá al suelo y permanecerá ahí; el fuego se elevará hacia el cielo.

Aquí en la Tierra, según el punto de vista de Aristóteles, si todas las cosas estuvieran en su estado natural, ya no se moverían. Se requiere una influencia externa para que las cosas se muevan, e incluso en tal caso el movimiento será meramente temporal. Podemos coger una piedra y arrojarla; se dará entonces un movimiento innatural o violento. Pero la piedra acabará por caer al suelo, quizá dando algunos rebotes, hasta volver a su estado natural de reposo sobre el suelo.

No estaba equivocado, esto es aplicable al menos para una amplia variedad de circunstancias. Si está usted sentado a la mesa con una taza de café delante, por sí misma la taza permanecerá ahí. Puede hacer que se mueva empujándola, pero cuando deje de empujar, la taza volverá a su estado de reposo. Podemos extrapolar esto, imaginaba Aristóteles, a modo de esquema básico del universo. El estado natural de los objetos es el de reposo, y el movimiento tan solo tiene lugar cuando algo los empuja fuera de su estado natural.



La representación no se ajusta tan bien con otros casos conocidos incluso en el tiempo de Aristóteles. Los antiguos griegos estaban muy familiarizados con flechas volando por el aire. La fuerza inicial le es aplicada por medio de la cuerda del arco, pero está claro que la flecha continúa desplazándose un tiempo después de haber partido del arco. ¿Por qué la flecha no se limita a caer al suelo? ¿Qué le impide volver de inmediato a su estado natural?

Esta es una cuestión a la que grandes mentes dieron vueltas durante siglos. La respuesta tardó en llegar, pero condujo finalmente al derrocamiento total de la perspectiva teleológica aristotélica acerca del universo, remplazada por una representación en la que los objetos no evolucionan hacia fines últimos, sino que obedecen a leyes que predicen lo que sucederá en el instante siguiente a partir de lo que está sucediendo ahora.

CONSERVACIÓN DEL MOMENTO

Juan Filópono, pensador alejandrino del siglo VI, dio un paso importante. Sugirió que la cuerda del arco transmitía a la flecha cierta cantidad, más tarde llamada «ímpetu», que mantiene a la flecha durante un tiempo en movimiento antes de terminar disipándose por completo. Una simple idea, tal vez, pero que suponía un paso importante, de pensar en términos de fines futuros a hacerlo en términos de propiedades existentes en el momento presente.

La idea de Filópono fue desarrollada por Ibn Sina (Avicena), polímata persa del siglo XI. Fue Ibn Sina quien dio el paso decisivo de sostener que el ímpetu no es tan solo temporal. Todo objeto tiene cierta cantidad de ímpetu (igual a cero en el caso de un objeto inmóvil, mayor que cero si el objeto está en movimiento), y esa cantidad permanece constante a menos que una fuerza cualquiera actúe sobre él.

En esta nueva representación, la razón por la que las piedras y las tazas de café dejan de moverse no es porque el reposo sea su estado natural, sino porque existen fuerzas —la fricción, la resistencia del aire— que van mermando gradualmente el ímpetu del cuerpo. En el vacío, sugirió Ibn Sina, no habría resistencia del aire, y un cuerpo en movimiento continuaría desplazándose perpetuamente a velocidad constante. Esto era poco menos que una locura especulativa hace mil años, pero hoy en día construimos habitualmente naves espaciales que se desplazan entre los planetas a una velocidad básicamente constante (salvo por el suave tirón de la gravedad). En el siglo XIV, el filósofo francés Juan Buridán propuso una fórmula matemática para el ímpetu, equiparándolo al producto del peso de un objeto por su velocidad.

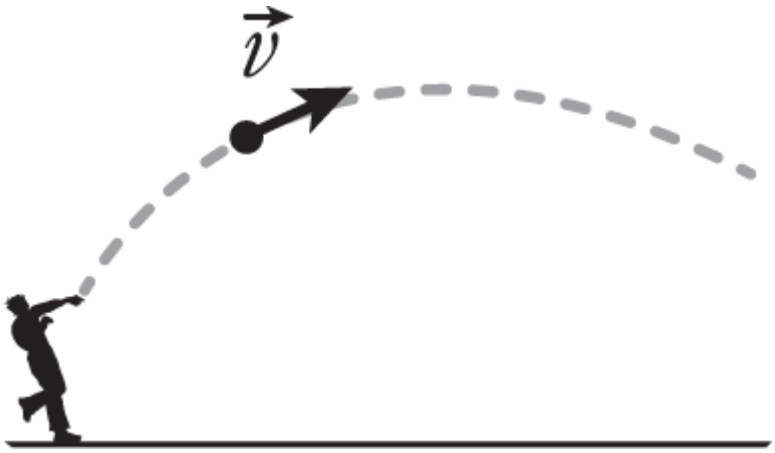
Lo que tenemos aquí es el nacimiento de una ley de la física: la conservación del momento. La idea somera de que había una «cantidad de movimiento» que se conservaba surgió antes de que nadie pudiera determinar con precisión cuál era esa cantidad. Se trata de un modo de avance estandarizado a lo largo de la historia en la física teórica: proponemos un concepto nuevo, trabajamos en él para caracterizarlo en términos cuantitativos, y luego tomamos esa formulación cuantitativa —una ecuación— y nos preguntamos cómo se comporta en relación con otros fenómenos que observamos en el mundo. Hoy sabemos que el momento o

cantidad de movimiento es el producto de masa por velocidad (al menos hasta que llega la relatividad y complica las cosas un poco).

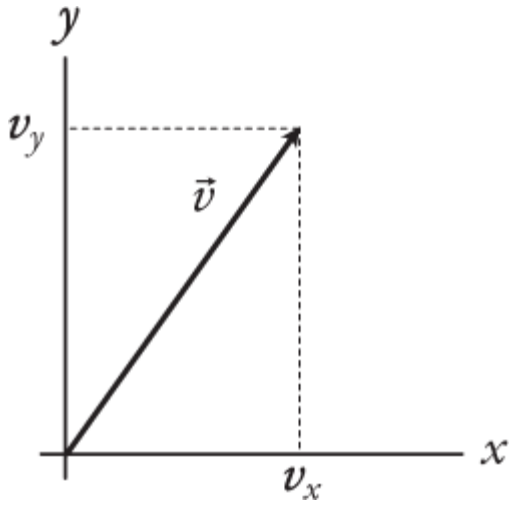
Un problema con la definición de Buridán de ímpetu como «peso por velocidad» es que «peso» no es una propiedad intrínseca de un objeto, ya que depende de la cantidad de gravedad que tire de él. Nuestro peso sería menor en la luna de lo que es en la Tierra, y pesaríamos menos también en una nave espacial que vagara entre los planetas. La masa, en cambio, sí es una propiedad intrínseca. Dicho en términos generales, la masa es la resistencia que un objeto opone a la aceleración. Se requiere mucha fuerza para acelerar un objeto de gran masa hasta una velocidad determinada, y tan solo una pequeña fuerza para acelerar hasta la misma velocidad un objeto de masa reducida.

De forma similar, celeridad y velocidad presentan una diferencia sutil. La rapidez o celeridad es un número, corresponde a metros por segundo; mientras que la velocidad es un vector, una cantidad compuesta de magnitud y dirección. De hecho, la magnitud del vector velocidad es justamente aquello a lo que llamamos «celeridad», solo que la velocidad apunta también a una dirección determinada. De modo que vamos con la misma celeridad si conducimos hacia el norte a 90 km/h que si conducimos hacia el sur a 90 km/h, pero la velocidad es diferente.

Señalamos los vectores dibujando una flechita sobre el símbolo pertinente; así, la velocidad de un objeto suele representarse mediante: . Muy a menudo nos interesa el tamaño, o magnitud, de un vector, que se escribe con el mismo símbolo pero sin la flecha: la magnitud de un vector es simplemente v.

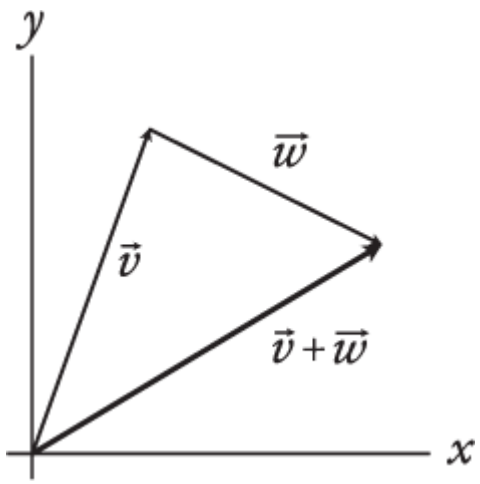


La notación en forma de flecha tiene sentido porque solemos representar una magnitud vectorial dibujando literalmente una flecha que apunta en la dirección del vector, y cuya longitud es proporcional a la magnitud del vector. Tenemos también la alternativa de representar un vector atendiendo a sus componentes: las contribuciones que recibe desde diferentes direcciones. Si estamos viajando exactamente hacia el norte, el componente de nuestra velocidad con respecto a la dirección este/oeste es cero.



Es fácil sumar vectores. Imagine simplemente que sitúa el comienzo del segundo vector al final del primero: definimos así como resultado un tercer vector desplazando hacia abajo el primero y luego el segundo. Si los dos vectores que estamos sumando apuntan (casi) en la misma dirección, el

vector resultante será (casi) tan largo como la suma de sus magnitudes respectivas, pero si apuntan en direcciones (casi) opuestas, el vector resultante puede ser mucho más corto.



Buridán y sus predecesores no pensaban en términos de vectores, que desarrollaron gradualmente cierto número de pensadores en el siglo XIX, entre ellos el matemático alemán August Ferdinand Möbius (famoso por la «cinta de Möbius»), el matemático irlandés William Rowan Hamilton, el polímata alemán Hermann Grassmann y el matemático inglés Oliver Heaviside. De modo que no cabe sorprenderse de que llevara su tiempo alcanzar la definición correcta de momento como cantidad de movimiento.

En nuestros días, el vector momento suele representarse con el símbolo . (Dado que la letra m está reservada para la masa, hemos adoptado este símbolo de la palabra latina que significa ‘momento’: petere). Con todo esto en mente, la expresión con que definimos el momento es la cosa más simple del mundo.

$$\vec{p} = m\vec{v} \qquad (1.1)$$

Nuestra primera ecuación oficial. El vector momento señala en la misma dirección que el vector velocidad, y sus magnitudes son proporcionales. La

proporcionalidad será un concepto fundamental para nosotros: significa que un cambio multiplicativo en una cantidad implica un cambio multiplicativo en la otra. Si duplicamos la velocidad, duplicamos el momento. El factor que relaciona ambas se llama «constante de proporcionalidad», aunque en algunas ecuaciones puede no ser constante en realidad. En este caso lo es: no es otra cosa que la masa del objeto.

La potencia de una ecuación incluso tan básica como esta se muestra evidente. No estamos diciendo que suceda que el momento de un objeto en particular resulte de multiplicar su masa por su velocidad; estamos diciendo que existe una relación universal entre el momento, la masa y la velocidad, relación que adopta siempre precisamente esta forma para todo objeto. Cuando lleguemos a la relatividad, algunas de las explícitas formulaciones de las ecuaciones que vamos a ver aquí van a experimentar ligeros retoques, pero los principios básicos son en amplia medida los mismos.

Una ecuación como esta no incluye en su seno ningún tipo de «causalidad»: expresa una relación rígida entre las cualidades involucradas, y tanto vale si se lee de izquierda a derecha como de derecha a izquierda. Podemos manejar la ecuación de cualquier forma que reproduzca la misma operación a un lado y a otro de la misma, por ejemplo, utilizando m como divisor, obteniendo: . De ahí que podamos decir: «Si conozco la velocidad de un objeto, puedo multiplicarla por su velocidad para obtener el momento»; como también: «Si conozco el momento de un objeto, lo dividiré por su masa para obtener su velocidad».

CHOQUES Y EMPUJONES

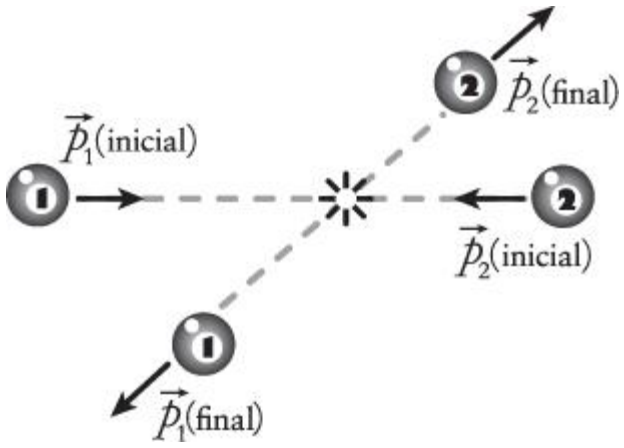
La potencia de la ley de conservación del momento va bastante más allá de la idea de que un objeto simple sin fuerzas que actúen sobre él seguirá desplazándose a una velocidad constante. Cuando una fuerza actúa sobre un objeto, por ejemplo, chocando contra otro objeto, el momento total del sistema entero se conserva.

Imaginemos que tenemos dos objetos moviéndose sin que ninguna fuerza actúe sobre ellos, como si se tratara de dos bolas de billar sobre una mesa carente de fricción. (Un disco sobre una mesa de aero-hockey sería levemente más realista, pero todo recurso a la «ausencia de fricción» resultará una idealización, por mucho que a los físicos les encante utilizar esta carta). Inicialmente se desplazarán siguiendo líneas rectas, pero terminarán por colisionar y desviarse para adoptar nuevas trayectorias rectas.

Designemos los momentos iniciales de la primera bola como 1 (inicial) y el de la segunda bola como 2 (inicial). Aplicando un patrón obvio, llamaremos al momento final de la primera bola 1 (final) y al de la segunda bola 2 (final). De tal manera, el enunciado de la conservación de momento en este caso será:

$$\vec{p}_1(\text{inicial}) + \vec{p}_2(\text{inicial}) = \vec{p}_1(\text{final}) + \vec{p}_2(\text{final}) \qquad (1.2)$$

Naturalmente, los momentos individuales cambian cuando las bolas rebotan una contra otra. Pero el momento total del sistema considerado como un todo se conserva.



El momento se conserva siempre, pero podemos perdonarles a nuestros predecesores que no lo advirtieran de inmediato. Volvamos a nuestra

aristotélica taza de café, que al principio estaba inmóvil, luego se movía un poco al empujarla y al final volvía al estado de reposo cuando usted dejaba de empujar. Parecería que el momento no se conserva, puesto que la taza va de una velocidad cero a una velocidad no cero. Pero, secretamente, la copa está reaccionando contra usted, y su propio cuerpo está ejerciendo un empuje contra la silla en la que está sentado, y la silla lo ejerce contra la propia Tierra. El cambio en el momento de la taza de café se compensa de forma exacta por el cambio en el momento en dirección opuesta que actúa en el sistema usted/silla/Tierra. Usted no lo advertirá, porque la masa de todo ese sistema es tan enormemente grande, que el cambio en su velocidad es inapreciablemente pequeño (aunque no absolutamente cero). Cuando usted deja de empujar y la taza vuelve a quedarse quieta, es porque la mesa sobre la que está ha ejercido un empuje de retroceso sobre ella, y esta vez el sistema mesa/Tierra recupera un pequeño momento, volviendo al estado inicial en que estaba antes de que usted empujara nada.

En la película Gravity, hay una escena en que Sandra Bullock y George Clooney están en peligro, flotando con sus trajes de cosmonauta fuera de una estación espacial. (Alerta de spoiler para este párrafo y el siguiente). Están sujetos a la estación por medio de una simple cuerda. Aunque la película es, por lo demás, de lo más amena, la física cojea un poco en esa escena. Lo que vemos es que Bullock tiene que agarrarse para salvar la vida mientras Clooney se ve impelido lejos de ella. En realidad, dado que están prácticamente en la misma órbita alrededor de la Tierra, no habría fuerzas que los impelieran lejos de la estación; un leve tirón de la cuerda los habría devuelto a la estación sanos y salvos. Pero en la película, ella tiene que soltarse de él para poder volver.

No tengo ningún reparo en que los cineastas fuercen un poco las leyes de la física en aras de una buena historia. Pero en este caso era innecesario. Lo único que había que hacer para tramar una escena igual de dramática era eliminar la cuerda sin más. Imaginemos que Bullock y Clooney hubieran estado agarrados el uno al otro, flotando y alejándose lentamente de la estación sin nada que los conectara a ella. La conservación del momento (cantidad de movimiento) les habría dicho exactamente lo que hacer. De haber permanecido juntos, se habrían alejado inexorablemente, a la deriva, y al final los dos habrían muerto. Pero había otra opción: uno de ellos podía

haber empujado al otro hacia el espacio. El momento total de ambos habría permanecido inalterable, pero uno de ellos habría podido volver hacia la estación, mientras el otro se habría alejado de ella todavía más deprisa. Clooney podría haberse sacrificado empujando a Bullock hacia la estación. Tanto Clooney como Bullock, o el uno o la otra, habrían podido salvarse empujando al otro a una inevitable sentencia. Pero eso habría sido otra película diferente.

MECÁNICA CLÁSICA

La conservación del momento sigue siendo un importante principio físico a día de hoy, pero el largo camino hacia su comprensión ha desempeñado un papel aún más importante al introducir una nueva forma de pensar la física. Lejos queda el mundo teleológico aristotélico con sus naturalezas intrínsecas, causas y efectos, y el movimiento que requería de un motor. Lo había sustituido un mundo de patrones, de leyes de la física. Después de algunas contribuciones importantes, a cargo de personajes como René Descartes o Galileo Galilei, el primer sistema completo de leyes físicas lo aportó Isaac Newton en 1687, mediante la teoría conocida como mecánica clásica.

Un aparte pedante, pero de importancia: los físicos modernos distinguen entre mecánica «clásica», que constituye un marco amplio, y mecánica «newtoniana», que es un modelo específico dentro de ese marco. La mecánica clásica dice que el mundo está hecho de cosas con valores definidos, mensurables, que obedecen a ecuaciones deterministas acerca del movimiento; esto contrasta con la mecánica cuántica. La mecánica newtoniana añade ideas específicas en torno al espacio y el tiempo absolutos; contrasta con la mecánica «relativista», que es clásica pero no newtoniana, y en ella el espacio y el tiempo se unifican. Hasta que empecemos a hablar explícitamente acerca de la relatividad, las ecuaciones que introduciremos para magnitudes como la energía y el momento serán newtonianas. Para complicar un punto más las cosas, hay teorías como la «mecánica lagrangiana» o la «mecánica hamiltoniana» que son

matemáticamente equivalentes a la mecánica newtoniana, pero que presentan conjuntos de términos y conceptos diferentes. Las mecánicas lagrangiana y hamiltoniana son ciertamente clásicas; si uno quiere caracterizarlas como newtonianas, es cuestión de gustos.

La mecánica clásica es una teoría de patrones, más que de naturalezas o de relaciones causa/efecto, porque no pregunta «¿cuál es el estado natural de un sistema?», o: «¿qué ha causado que el sistema se mueva de este modo?». Todo lo que pregunta es: «¿qué está haciendo el sistema en este momento particular?». Y a partir de ahí, realiza una predicción precisa acerca de lo que el sistema estará haciendo en cualquier otro momento. Puede ser un momento del pasado, no solo del futuro. Volviendo a la ecuación (1.2), la conservación del momento lineal para dos partículas, las consecuencias funcionan tanto hacia atrás en el tiempo como hacia delante. Si conocemos el momento total final, sabemos que era el mismo en instantes anteriores del tiempo.

Esto constituye un ejemplo de una ley de la conservación diferente, mucho más impresionante: la conservación de la información. Este principio estaba implícito en las leyes de Newton de la mecánica clásica, pero no se puso de relieve hasta la obra del matemático francés Pierre-Simon Laplace, hacia 1814. El estado de un sistema clásico en cualquier momento del tiempo queda especificado por las posiciones y velocidades de cada parte del sistema; por ejemplo, de todos los planetas del sistema solar. Esta cantidad de información, señaló Laplace, se conserva a lo largo del tiempo. A partir del estado de cualquier momento dado, podemos predecir el estado de cualquier otro momento, futuro o pasado. O como mínimo podríamos, si contáramos con un perfecto conocimiento de la información y con unas capacidades de cálculo arbitrariamente adecuadas. Laplace imaginó un «vasto intelecto» con tales capacidades, pero algunos comentaristas posteriores llamaron a este ser hipotético el demonio de Laplace. La cuestión relevante del experimento mental del demonio de Laplace no es que una persona pudiera realmente tener tal información y realizar tales predicciones, en un sentido práctico, ni que debamos aspirar a ello. Es completamente imposible que ningún ser real pudiera conocer las posiciones y velocidades de cada átomo que se encierra en un grano de arena, cuánto menos en el universo entero. Pero el propio universo posee

esta información, y las leyes de la mecánica clásica predicen que tal información se conserva a lo largo del tiempo.

CONSERVACIÓN DE LA ENERGÍA

Tenemos luego la conservación de la energía, uno de los distintivos más conocidos de la mecánica clásica, que constituye además un interesante ejemplo del desarrollo de las ideas filosóficas. A diferencia del momento, que es un vector, la energía de un objeto es simplemente un número, una cantidad con magnitud pero sin dirección. (A los números se les denomina a veces «escalares», cuando queremos distinguirlos específicamente de los vectores o de cantidades más complicadas). La energía se presenta bajo cierto número de formas, pero hay un tipo de energía —la energía cinética, la energía asociada al movimiento— que está relacionada con el momento. La fórmula[1] para expresar la energía cinética de un objeto de masa m y velocidad v es:

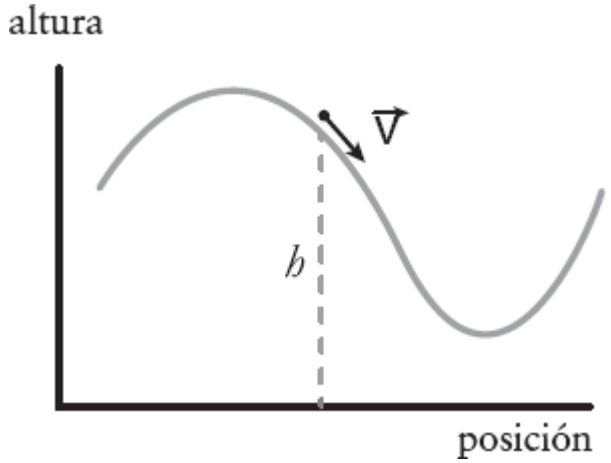
$$E_{\text{cinética}} = \frac{1}{2}mv^2 \qquad (1.3)$$

Tanto el momento como la energía se conservan en la mecánica clásica, pero no así la energía cinética como tal, puesto que puede convertirse en (o crearse a partir de) otras formas de energía. Cuando disparamos una flecha con un arco, la energía que se encerraba en la tensión de la cuerda se convierte en la energía cinética de la flecha.

En circunstancias simples, podemos rastrear directamente cómo la energía se transforma de una forma en otra. A los físicos les gusta pensar en una pelota rodando por una colina, prescindiendo de la fricción y de la resistencia del aire. En tal situación, existe una energía potencial asociada a la elevación de la pelota sobre la colina, además de la energía cinética. La fórmula de la energía potencial de una pelota situada a una altura h es:

$$E_{\text{potencial}} = mgh \qquad (1.4)$$

Aquí, m es la masa de la pelota y g es la aceleración debida a la gravedad cerca de la superficie terrestre. (Tiene libertad para dar a g el valor apropiado para la aceleración si realiza el experimento en otro planeta). En valores numéricos, g = 9,8 metros por segundo al cuadrado: si dejamos caer un objeto (prescindiendo de la resistencia del aire), dicho objeto incrementa su velocidad de caída en 9,8 metros/segundo por cada segundo. Tal sería la aceleración si no hubiera colina alguna.

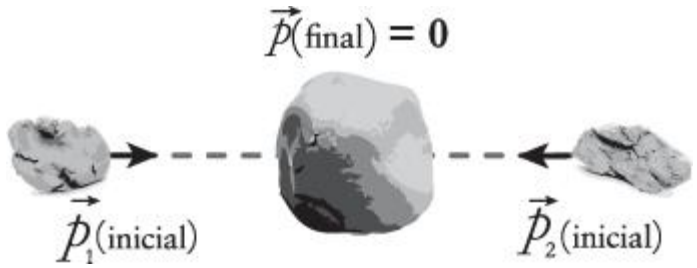


Rodando por una colina sin fricción, la energía total Ecinética + Epotencial permanece constante, pero ambas formas de energía se convierten la una en la otra. Si la pelota parte estando inmóvil en lo alto de una pendiente, comenzará rodando hacia abajo, y en cada instante su velocidad será exactamente la requerida para producir una cantidad de energía cinética igual a la energía potencial que se pierda.

Es fácil ver cómo la energía potencial y la cinética se convierten la una en la otra, pero otras formas de energía son más difíciles de discernir. Hemos puesto antes el ejemplo de dos bolas de billar chocando y desviándose por efecto de los choques, en concreto de dos «bolas de billar de físicos», que se desplazan por una superficie totalmente carente de fricción y que no

producen sonido alguno ni calor cuando colisionan. En este caso, tanto el momento como la energía cinética se conservan. Llegados a este punto, es posible que tenga usted flashbacks de alguna clase de física del instituto o del colegio. Cuando ambas cantidades se conservan, estamos hablando de choques elásticos, ya que los objetos involucrados simplemente rebotan separándose el uno del otro.

Pero también hay choques inelásticos, en los que el momento se conserva, pero la energía cinética se transforma en alguna otra forma. Imaginemos que, en lugar de bolas de billar, hacemos chocar dos terrones de barro, cada uno de los cuales parte con una cantidad de momento opuesta: 1 (inicial) = − 2 (inicial). En lugar de rebotar y separarse, los dos terrones se deforman un poco y se quedan pegados. El momento se conserva, pero no así la energía cinética tal cual, ya que se ha convertido en calor y tensión en el interior del terrón de barro.



En un principio, los pensadores, incluido el propio Newton, no habían advertido que el momento y la energía eran dos cosas separadas; pensaban en términos de una simple «cantidad de movimiento». En el contexto de la mecánica newtoniana, en la que el movimiento rectilíneo constante en los cuerpos no sometidos a otras fuerzas constituye un principio fundamental, el momento es algo cuya definición resulta obvia. La energía cinética era algo más sutil, aunque pensadores como Gottfried Wilhelm Leibniz (el rival de Newton en la invención del cálculo) sostuvieron que la cantidad central en el estudio del movimiento era la «vis viva», que él definió como mv^2.

Finalmente aclaró la situación la filósofa y física francesa Émilie du Châtelet, quien había traducido la obra de Newton al francés. Valoraba la noción de conservación del momento, pero sostenía que la energía

conservada representaba una magnitud diferente. Como prueba, realizó un experimento ideado por el físico holandés Willem ‘s Gravesande, consistente en dejar caer unas pesadas bolas sobre arcilla. El momento de la bola se transfiere al barro en su totalidad, pero su energía cinética se emplea en crear una depresión en la superficie. Se constató que la cantidad de arcilla desplazada era proporcional al el cuadrado de la velocidad en el instante del impacto, exactamente lo expresado por la fórmula de la energía cinética.

Quizás haya oído hablar de la ley de la «conservación de la masa», pero ha dejado de ser estrictamente cierta desde la aparición en escena de la teoría de la relatividad. Según la relatividad, la energía y el momento se conservan (aunque las ecuaciones que los expresan son ligeramente diferentes de las que hemos escrito antes), pero la masa no es más que una forma particular de energía. Tal es el significado de la famosa ecuación de Einstein: la energía de un objeto en reposo (es decir, con energía cinética igual a cero) es simplemente el producto de su masa por la velocidad de la luz al cuadrado.

$$E_{\text{reposo}} = mc^2 \qquad (1.5)$$

En circunstancias normales, postular conservación de la masa resulta una aproximación bastante buena, pero en física de partículas, en la que las partículas se mueven rutinariamente a una velocidad cercana a la de la luz, deja de ser útil, y debemos limitarnos a pensar en conservación de energía.[2]

¿POR QUÉ HAY LEYES DE LA CONSERVACIÓN?

A los científicos nos encanta preguntar «¿por qué?». Queremos saber por qué las manzanas caen de los árboles, por qué el café y la leche se mezclan, pero nunca se separan, por qué un fuego se apaga si lo privamos de

oxígeno… Pero es una cuestión peliaguda, tenemos que enfrentarnos a la posibilidad de que nuestros «¿por qué?» toquen fondo al encontrarse con las últimas respuestas, más allá de las cuales no parece haber nada más que: «pues porque las cosas son así».

Afortunadamente, en lo que se refiere a las leyes de conservación quizá podamos sacar algo más en claro. No fue hasta comienzos del siglo XX cuando la matemática alemana Emmy Noether[3] demostró finalmente un notable teorema que relacionaba las leyes de conservación con simetrías de las leyes de la naturaleza. La idea básica es sencilla: una simetría es una transformación que puede hacérsele a un sistema de tal forma que deje sus características esenciales sin cambio. Un círculo tiene una simetría por la que podemos hacerlo rotar en cualquier ángulo alrededor de su centro; un cuadrado, en cambio, es simétrico sometido a rotaciones de 90 grados, o múltiplos de 90.

El teorema de Noether postula que toda transformación simétrica, regular y continua de un sistema conlleva la conservación de cierta cantidad. Las leyes de la física, por ejemplo, en su conjunto, son simétricas en relación con cambios en el espacio (podríamos trasladar el sistema entero a otro lugar) y también con cambios en el tiempo (podemos realizar un experimento, esperar un rato y repetirlo). En cualquiera de ambos casos obtendríamos el mismo resultado antes y después del cambio. El teorema de Noether relaciona estas simetrías con las leyes de conservación que ya conocemos: la invariancia con respecto a los cambios espaciales nos lleva a la conservación del momento y la invariancia con respecto a los cambios temporales nos conduce a la conservación de la energía. El número de las diferentes simetrías sale correcto. En cuanto al tiempo, tenemos una sola dimensión, y en consecuencia una sola cantidad conservada: la energía. Pero tenemos tres dimensiones en cuanto al espacio, por lo que podemos trasladar el sistema en cualquiera de sus tres direcciones correspondientes por separado. Por eso el momento es un vector, del cual podemos concluir que tiene tres componentes, uno por cada dirección en el espacio. Así, hay también una simetría correspondiente a cada una, por la que podemos rotar el sistema alrededor de cualquiera de los tres ejes independientes. Ello da lugar a la conservación de otra cantidad: el momento angular.

SIMETRÍA	CANTIDAD CONSERVADA
Cambios temporales	Energía
Cambios espaciales	Momento
Rotaciones	Momento angular

Estas simetrías —cambios en el espacio, cambios en el tiempo, rotaciones — son todas ellas simetrías del espacio-tiempo, ya que implican transformaciones del sistema en el espacio y en el tiempo. En física de las partículas y en teoría cuántica de campos tenemos simetrías internas en que «rotan» diferentes partes de un campo cuántico a otro. Estas simetrías internas son las responsables de la conservación de la carga eléctrica y de cierto número de otras propiedades de las partículas.

Descubrimos una sutileza si recordamos que las leyes de la física muestran una simetría, pero que tal simetría puede quebrantarse por la situación efectiva en que nos encontremos nosotros mismos. Nuestro universo, por ejemplo, se expande. Las galaxias remotas se alejan gradualmente unas de otras a medida que pasa el tiempo. En consecuencia, cobra sentido pensar que la energía no se conserva en un universo en expansión. La configuración de nuestro universo no es invariable a los cambios en el tiempo; antes las cosas estaban más juntas, y en futuro estarán aún más separadas. Si tomamos la energía contenida bajo todas las diferentes formas de materia que conocemos (radiación, materia corriente, materia oscura, energía oscura, su propia materia) y lo sumamos todo, obtenemos un número que no es constante en el tiempo. Hay formas de intentar fijarlo definiendo una energía en la curvatura del espacio-tiempo mismo, pero no son completamente satisfactorias. No hay nada malo en definir «la energía total contenida en una región del espacio» como «la suma de las energías de todas las cosas incluidas en esa región», y admitir que tal número cambia con el tiempo.

Todo lo cual es tanto como decir: las leyes de conservación son delicadas. Mejor no perder la cabeza. Probablemente sea esta una lección de amplia aplicabilidad mientras seguimos profundizando en las más grandes ideas del universo.

LA FILOSOFÍA DE LA VACA ESFÉRICA

Las leyes de conservación, obviamente, son tan importantes conceptualmente, como útiles en la práctica. Pero hay otro motivo por el que representan un punto de partida apropiado, en particular la conservación del momento, para seguir adelante con la exploración del mundo físico: ejemplifican un principio metodológico básico, la filosofía de la vaca esférica.

La expresión procede de un chiste que a los físicos les gusta contar acerca de sí mismos. Había un granjero de vacas lecheras que tenía problemas con el rendimiento de su granja y decidió pedir ayuda a un científico de la universidad. Por motivos desconocidos, consultaron con un físico teórico. Tras entregarse a complicados cálculos, el físico volvió con un montón impresionante de ecuaciones. «Creo que he solucionado su problema», dijo el físico. «Ah, ¿sí? Cuénteme», replicó el ganadero, expectante. «Bien, en primer lugar supongamos una vaca esférica…».

La gracia del chiste, por si no resulta obvio de inmediato, está en que las vacas, no solo no son esféricas, sino que su naturaleza no esférica es crucial para lo que significa ser una vaca. Una vaca esférica no sería en absoluto una vaca. La suposición del físico puede simplificar los cálculos requeridos por él, pero al hacerlo nos aleja por completo del ámbito de conceptos que de verdad podrían ser útiles para el granjero.

El chiste es famoso no solo porque pueda ser o no desternillante —nunca nadie ha reivindicado tal cosa—, sino porque el equivalente de «supongamos una vaca esférica» realmente funciona en física, y funciona increíblemente bien. Representa un ejemplo de un principio general, que podríamos describir como: simplificar un problema difícil hasta reducirlo a su expresión más sencilla, ignorando todas las complicaciones que se pueda. Una vez obtenida la respuesta al problema sencillo, restituirle las complicaciones y calcular en qué medida estas afectan a la solución del problema simplificado.

Mediante este método de pensamiento descubrimos la conservación del momento. Aristóteles no estaba equivocado: las tazas de café no se mueven por sí mismas, y si las empujamos unos segundos y luego paramos, inmediatamente vuelven a su estado de reposo. Pero tampoco Ibn Sina se equivocaba: la copa vuelve al estado de reposo a causa de la fricción, no de

su naturaleza intrínseca. Si ignoramos la fricción e imaginamos una flecha volando a través de un espacio vacío, continuará desplazándose a velocidad constante. Este es un punto de partida para un análisis físico útil; más tarde siempre podemos añadirle las complicaciones de la fricción.

El maestro en este tipo de razonamiento fue Galileo. Tenía un don especial para desentrañar qué cosas eran esenciales y cuáles podrían ignorarse en un primer momento. Aristóteles había afirmado que los objetos más pesados caían más deprisa que los ligeros. Tampoco esta vez se equivocaba: deje caer un libro y una hoja de papel al mismo tiempo, y compruebe lo que revela el resultado de su experimento. Pero Galileo argumentó que si no hubiera resistencia del aire, caerían con la misma velocidad. Se trataba de un experimento que nadie pudo llevar a la práctica antes siglos más tarde, pero Galileo fue capaz de construir ingeniosos montajes experimentales con los que pudo establecer el principio subyacente.

Veremos aplicar la filosofía de la vaca esférica una y otra vez. Pero por útil que sea, no es un método universal de descubrimiento de la verdad. No funciona tan bien con las granjas de vacas lecheras, por ejemplo. En muchos sistemas complejos, como los que con frecuencia encontramos en el mundo macroscópico a escala humana, hay aspectos diferentes que implícitamente interactúan en cada situación entre sí de un modo determinante, hasta el punto de que no podemos, simplemente, ignorar uno en cada ocasión y corregir sus efectos más tarde. En ámbitos como la biología o la economía, a menudo todas las cosas dependen de todas las demás.

La razón por la que la física parece tan difícil es porque en realidad es bastante sencilla, en comparación con otras ciencias. En física (al menos en ciertas partes) disponemos de este recurso milagroso de separar varios componentes del sistema sobre el que nos estamos cuestionando, resolver un problema mucho más simple, y volver a ensamblarlo todo al final. Esto nos hace la vida mucho más fácil, en comparación con lo que podría ser. Como resultado, los físicos somos capaces de descubrir características asombrosas y antiintuitivas, desde la mecánica cuántica hasta la relatividad, pasando por el Big Bang. Jamás habríamos sido capaces de suponer estas cosas mediante la simple inteligencia: teníamos que inventarlas intentando

hacer encajar los datos experimentales. Pero precisamente por ser tan antiintuitivas, pueden resultar difíciles de entender, al menos a primera vista. Si el panorama que se despliega ante nosotros nos parece tan poco familiar, es porque estamos mirando verdaderamente lejos.

1 ¿Por qué ½? Hay una razón, pero se requiere del cálculo para explicarla, y aún no hemos llegado a ello. Para futuras referencias: cuando empujamos un objeto, la energía acumulada es la integral de la fuerza según cambia la distancia. La segunda ley de Newton equipara la fuerza al producto de la masa por la aceleración. La aceleración es la derivada de la velocidad respecto del tiempo. De modo que la integral de la aceleración sobre la distancia es igual a la integral de la velocidad a medida que la propia velocidad cambia. Es decir, . Quizá nada de esto tengo sentido ahora, pero no tiene más que esperar.

2 Oirá también hablar a veces de «masa relativista», que aumenta con la velocidad. Esto es confuso e innecesario; es mejor considerar simplemente la masa de un objeto como una cantidad fija y dejar que la energía dependa de la velocidad.

3 Los apellidos que en alemán llevan una «o» con signo diacrítico (umlaut): «ö», suelen escribirse en inglés como «oe». Así, vemos en ocasiones escrito el apellido de Erwin Schrödinger como «Schroedinger». De ahí que podríamos pensar que el apellido de Emmy Noether fuera en realidad «Nöther»; pero no es el caso, ella lo escribía así, «Noether». Es un caso similar al de Johann Wolfgang von Goethe.

DOS

CAMBIO

La física es posible porque el mundo muestra cierta cantidad de continuidad y predictibilidad. Pero sería en verdad un mundo muy aburrido si todas las cosas que nos rodean estuvieran siempre igual, sin cambiar nunca en nada. Por suerte no es así. Los planetas y las estrellas se desplazan por el espacio, la gente va a trabajar o coge un avión para irse de vacaciones, los átomos individuales se mueven zumbando por el interior de su cuerpo y por todas partes. Necesitamos poner toda nuestra atención para entender este comportamiento lo mejor posible.

En la física clásica, para describir el cambio se recurre a un marco conceptual específico al que llamamos paradigma laplaciano, por Pierre-Simon Laplace. El experimento mental del demonio de Laplace servía para ilustrar la conservación de la información: los datos necesarios para predecir el futuro o retroinferir el pasado están presentes en cualquier momento de la historia de un sistema aislado. Esto sugiere el siguiente paradigma para entender cómo cambian las cosas:

•Especificar el estado completo del sistema en un momento del tiempo.

•Utilizar las leyes de la física para calcular qué estado habrá justo un momento después (o antes).

•Utilizar de nuevo las leyes de la física para avanzar de ese nuevo momento al siguiente.

•Repetir el proceso.

Este algoritmo nos permite trazar la historia entera del sistema, pasado, presente y futuro. No hay ninguna referencia a hipotéticas naturalezas, propósitos o metas hacia los que el sistema pudiera aspirar. Es simplemente una cosa después de otra, y la siguiente cosa está determinada por la cosa actual y las leyes de la física.

Pero ¿qué diantre queremos decir con «justo un momento después»? ¿Qué se supone que denota aquí un «momento»? ¿Un segundo? ¿Una millonésima de segundo? Suena un poco arbitrario.

Lo que pretendemos, presumiblemente, es descomponer el tiempo hasta su unidad más pequeña posible. Pero el tiempo parece discurrir uniformemente, tanto desde la óptica de la experiencia común, como de la perspectiva de la física moderna. La continuidad del tiempo era reconocida con desconcierto ya por parte de los pensadores antiguos, desde Zenón de Elea (y sus famosas aporías) hasta Arquímedes. El desconcierto no llegó a superarse realmente hasta que Isaac Newton y Gottfried Wilhelm Leibniz descubrieron independientemente el cálculo, la técnica matemática para trabajar con cantidades infinitesimales. De modo que deje a un lado todo miedo que hubiera podido adquirir en el instituto o el colegio: vamos a tener que aprender algo de cálculo en este capítulo.

La buena noticia es que los conceptos básicos son más fáciles de captar de lo que le hayan podido inducir a pensar. De hecho, solo hay dos: utilizar «derivadas» para calcular la razón de cambio de algo e «integrales» para calcular la cantidad total de cambio.

(No hay ninguna mala noticia).

PLANETAS Y FUERZAS

Con el fin de abrirnos paso para comprender el paradigma laplaciano que describe el cambio, es útil considerar la alternativa. Un ejemplo nos lo proporciona el movimiento de los planetas en el sistema solar, que históricamente ha sido una fructífera fuente de ideas para el conocimiento de la física.

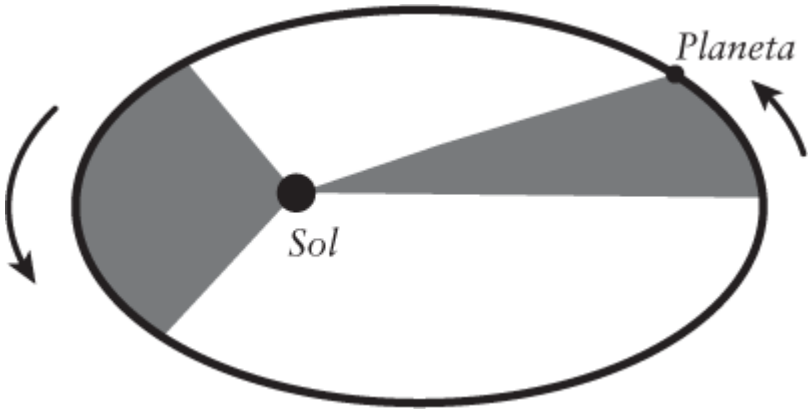
La mayoría de nosotros conoce la historia de Ptolomeo y Copérnico. Ptolomeo, astrónomo alejandrino del siglo II, desarrolló un modelo geocéntrico (con la Tierra en el centro) del sistema solar, que permaneció vigente durante más de un milenio. Copérnico, astrónomo polaco del siglo XVI, propuso un modelo alternativo, el modelo heliocéntrico (con el sol en el centro). Esto inquietó a unas cuantas personas, que preferían pensar que ocupaban el centro de todo. Pero tanto Ptolomeo como Copérnico basaban sus modelos en la geometría de círculos y, como resultado, ambos necesitaban ser retorcidamente complicados para poder hacer encajar en el modelo las observaciones astronómicas conocidas. Si querían representar las órbitas planetarias sirviéndose únicamente de círculos, tenían que desviar individualmente cada órbita, con lo que dejaban de compartir un centro común, y los planetas no se movían ya en sus respectivos círculos principales, sino en epiciclos circulares más pequeños centrados en los círculos principales.

La situación se simplificó en gran medida en el siglo XVII gracias al astrónomo alemán Johannes Kepler. Aceptó el desplazamiento de la Tierra propuesto por Copérnico, en favor de que el sol ocupara el centro del sistema solar, pero la verdadera innovación de Kepler fue sugerir que, por estupendos que sin duda fueran los círculos, había otras figuras que cabía tener en cuenta a la hora de describir las órbitas planetarias. Planteó la elipse, y de golpe quedó eliminada por completo la necesidad de los epiciclos y sus complicaciones asociadas. A partir de los datos que habían sido laboriosamente recopilados por su mentor, Tycho Brahe, Kepler propuso tres leyes para el movimiento de los planetas.

1. Los planetas se desplazan siguiendo elipses, siendo el sol uno de los focos de la elipse.

2. La órbita de un planeta barre áreas iguales en tiempos iguales. En consecuencia, los planetas se desplazan más rápidamente cuando están cerca del sol y más lentamente cuando se alejan de este.

3. Cuanto mayor es la órbita, más largo es el periodo orbital. En particular, el cuadrado del periodo es proporcional al cubo del eje mayor de la elipse. Esto hace que las órbitas de los diferentes planetas estén relacionadas entre sí.



Las leyes de Kepler representaron un enrome paso adelante en nuestra comprensión de la dinámica de los planetas. Pero, como sucede con todo avance, surgieron nuevas preguntas. ¿Por qué las órbitas de los planetas obedecen a estas leyes? ¿Es esta una pregunta sensata siquiera?

Esta pregunta y otras relacionadas preocuparon a algunas de las mejores mentes del mundo hacia el final del siglo XVII, en los albores de la Edad de la Razón. Inspirados por la obra de Galileo y Descartes, los filósofos de la

naturaleza centraron su mirada en el terreno de la mecánica, el estudio general del movimiento y de cómo se produce. A partir del nuevo conocimiento de que el momento se conserva y que el movimiento constante rectilíneo es un estado natural de las cosas, era importante comprender por qué los objetos tendrían que desviarse nunca de las líneas rectas. La respuesta obvia era que sobre ellos actuaba algún tipo de fuerza, pero al principio era algo oscuro determinar qué era lo que había que contar como una fuerza y cómo operaban realmente las fuerzas. No estábamos tan lejos del concepto aristotélico del movimiento violento o innatural, pero con una idea diferente acerca de lo que constituye un movimiento natural: para Aristóteles era natural estar en reposo, mientras que para la mecánica clásica lo natural es moverse a una velocidad constante cualquiera.

Un importante paso adelante lo dio el físico holandés Christiaan Huygens. Imaginemos que atamos una piedra al extremo de una cuerda y la hacemos girar en círculo. Notaremos que la piedra tira de la cuerda, y nosotros tendremos que tirar hacia nosotros para contrarrestar esa fuerza y mantener el movimiento circular. Huygens obtuvo una fórmula para calcular la cantidad de «fuerza centrípeta» que necesitábamos en nuestro esfuerzo por lograr mantener dicho movimiento. (Propuso también la teoría ondulatoria de la luz, inventó el reloj de péndulo y descubrió Titán, uno de los satélites de Saturno. Fue una época de grandes emociones). No es lo mismo que sucede con los planetas que se mueven alrededor del sol, pero las similitudes básicas son evidentes.

En 1660 se fundó en Inglaterra la Royal Society, que pronto contó con pensadores tales como el científico experimental Robert Hooke, el arquitecto Christopher Wren y el joven astrónomo Edmond Haley, los cuales se hicieron amigos. Hooke había sugerido en una conferencia impartida en la sociedad que el movimiento de los planetas podía explicarse por la acción de una fuerza gravitacional, que se extendía desde el sol y decrecía con la distancia, y que desviaba suavemente la trayectoria del planeta haciendo que no fuera en línea recta. Pero Hooke no se sentía totalmente a sus anchas con las matemáticas, por lo que no era capaz de elaborar a partir de esta idea un sistema integral. Pasaba tiempo en los cafés (una popular novedad de la época), discutiendo acerca de la cuestión con Wren y Halley. Para 1684, los amigos habían comprendido que la fórmula

de Huygens de la fuerza centrípeta podía usarse para llegar a la tercera ley de Kepler, si la fuerza de gravedad del sol obedecía a una ley de la inversa del cuadrado: la atracción gravitatoria decrece en proporción al cuadrado de la distancia a un planeta. Un planeta que estuviera alejado al triple de distancia, percibiría una novena parte de la fuerza. Lo que no fueron capaces de demostrar era que una órbita sometida a este tipo de influjo tomaría realmente la forma de una elipse.

NEWTON Y SUS LEYES

Wren, que quizá tenía más dinero en metálico que los otros dos gracias al diseño de la catedral de Saint Paul y otras edificaciones, ofreció un premio a quienquiera que fuera capaz de obtener la forma de las órbitas planetarias en correspondencia con tal ley de la inversa del cuadrado. Ya por aquella época todo el mundo sabía que Isaac Newton era el más inteligente de todos, aunque se pasara la mayor parte del tiempo sin salir de Cambridge. Finalmente, Halley reunió el valor necesario para hacerle una visita. Newton lo recibió con amabilidad, y Halley le planteó la cuestión de la forma de las órbitas ligada a una fuerza inversa del cuadrado. Newton repuso al instante que sería una elipse. Pillado un poco por sorpresa, Halley insistió, preguntándole cómo podía estar tan seguro. «Bueno», replicó Newton, «ya lo he calculado».

Sigue sin estar muy claro cómo llevó exactamente a cabo Newton el cálculo original. Lo que sabemos es que Halley le instó a que escribiera algo al respecto, propuesta que se tradujo en un breve artículo que Halley presentó ante la Royal Society. A aquella primera publicación le faltaban algunos detalles, por lo que Halley urgió al gran hombre a que la desarrollara un poco. Una vez se aplicó a la tarea, Newton ya no pudo parar. Dieciocho meses más tarde, había escrito Philosophiæ naturalis principia mathematica (Principios matemáticos de la filosofía natural), comúnmente abreviado como «los Principia», de la que podría sostenerse que sea la obra más influyente de la historia intelectual moderna.

Los Principia no solo establecieron el sistema de la mecánica clásica que se mantuvo inalterable hasta el siglo XX, sino que propuso también la ley de gravitación completa, derivada de las leyes de Kepler a partir de principios básicos, y presentó los primeros indicios de lo que ahora conocemos como cálculo. Newton había avanzado mucho en el desarrollo del cálculo, pero era reacio a desplegarlo plenamente en los Principia, ya que las técnicas eran nuevas y susceptibles de generar controversia. Como consecuencia, terminó en una amarga disputa con Leibniz, quien desarrollo aquellas ideas de forma independiente y cuya notación acabaría pasando al uso común. (Newton terminó también en amarga disputa con Hooke en torno a la ley de la inversa del cuadrado. Terminó en un montón de amargas disputas).

La mecánica clásica es un sistema, no una mera teoría sobre de un fenómeno físico en particular. Constituye una forma integral de pensamiento acerca de cómo funciona el mundo físico, forma de pensamiento que fue casi universalmente aceptada como verdadera hasta la aparición de la mecánica cuántica. Hemos visto el gran número de personas de elevado grado de inteligencia que se debatieron a lo largo de siglos para captar una noción de momento, fuerza y movimiento. Newton pensó acerca de todo ello con detenimiento y presentó lo que simplemente parecía ser la respuesta acertada. Como ha señalado el cosmólogo contemporáneo Rocky Kolb: «Para comparar el logro de Newton con el del primer vuelo tripulado a motor, tendríamos que imaginar a los hermanos Wright, Orville y Wilbur, elevándose sobre la arena de Kitty Hawk el 17 de diciembre de 1903, tras los controles de un moderno avión de pasajeros en vuelo hacia Nueva York».[1]

Entonces, ¿qué dijo Newton en realidad? Ahondaremos en la mecánica clásica en el capítulo siguiente, pero en lo que respecta al movimiento de los planetas, hay dos ideas fundamentales.

En primer lugar, si el movimiento natural de los objetos consiste en continuar en línea recta a velocidad constante, las desviaciones de dicho movimiento pueden caracterizarse por la aceleración de ese cuerpo, la razón a la que cambia la velocidad. Como la velocidad misma, la aceleración será un vector, pues podemos imaginar acelerar en direcciones diferentes. La segunda ley del movimiento de Newton, la ecuación más famosa de la física

clásica, determina que la aceleración de un objeto es proporcional a la fuerza neta que actúa sobre él, con la masa del objeto como constante de proporcionalidad:

$$\vec{F} = m\vec{a} \qquad (2.1)$$

La primera ley dice que la velocidad se mantendrá constante cuando no se dé ninguna fuerza, y la tercera ley postula que los cuerpos actúan uno sobre otro con fuerzas iguales y opuestas. Con frecuencia habrá más de una fuerza actuando a la vez sobre un objeto, en cuyo caso no tendremos más que sumar las contribuciones individuales para obtener la fuerza total, responsable de la aceleración. (Dado que los vectores tienen dirección además de magnitud, dos fuerzas grandes pueden dar como resultado al sumarse una fuerza neta pequeña, si apuntan a direcciones opuestas o casi opuestas). Podemos dividir ambos miembros de (2.1) entre m para obtener la aceleración en términos de fuerza: .

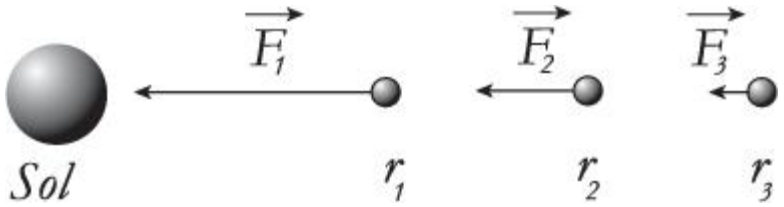
La segunda idea es la ley de gravitación universal de Newton, versión final de la ley de la inversa del cuadrado de la que Hooke y otros habían hablado. Lo que era incontestablemente propio de Newton era la noción de que la ley fuera realmente universal: explica desde cómo las manzanas caen de los árboles, hasta cómo los planetas se mueven alrededor del sol. Es algo que ahora damos por hecho, pero en aquel tiempo relacionar el movimiento de los planetas con los hechos cotidianos del huerto de tu casa suponía un salto portentoso.

Tomemos dos cuerpos celestes de masa m1 y m2, separados por una distancia r. Sea r el «vector unidad» —un vector de longitud 1, cualesquiera que sean las unidades que usemos para medir la distancia—, cuya dirección es del objeto 2 al objeto 1. Newton dice que la fuerza que actúa sobre el objeto 2, debida a la atracción gravitatoria del objeto 1, se expresa:

$$\vec{F} = G\frac{m_1 m_2}{r^2}\vec{e}_r \qquad (2.2)$$

El número G es una constante de la naturaleza, conocida ahora como la constante de gravitación de Newton, que nos dice cómo es de intensa la fuerza de la gravedad. Desde el punto de vista del objeto 2, el objeto 1 es la fuente de gravedad —la propiedad física que crea la fuerza—, y viceversa.

Esta ecuación tiene un aspecto algo más elaborado que las anteriores, pero si nos detenemos a pensarlo, una vez más advertimos una relación de proporcionalidad entre dos vectores: la fuerza ejercida sobre el objeto 2 por parte del objeto 1, y un vector r como unidad de longitud que apunta de 2 a 1. La complejidad de la notación simbólica refleja simplemente el hecho de que el factor de proporcionalidad afecta tanto a la constante de Newton como las dos masas, y que luego lo dividimos por el cuadrado de la distancia. Ello asigna un valor numérico preciso a la cantidad de atracción que el sol ejerce sobre los planetas, la cual es sustancial cuando nos acercamos al sol y va disminuyendo a medida que nos alejamos de él.



A partir de estas dos simples pautas, la segunda ley del movimiento (2.1) y la ley de gravitación universal (2.2), Newton fue capaz de reformular todas las leyes de Kepler, y mucho más. Mostró, por ejemplo, cómo la fuerza de gravitación debida a un cuerpo esférico sería la misma si esa misma cantidad de masa estuviera concentrada en un solo punto en el centro. De modo que es correcto tratar al sol y los planetas como si fueran simples puntos en lugar de objetos sólidos, al menos en la aproximación por la que los representamos como esferas perfectas. Y lo que es tan importante como esto, podemos ir más allá de la imagen ideal por la que considerábamos cada uno de los planetas en órbitas aisladas alrededor del sol: con las pautas de Newton en la mano, podemos preguntarnos, por ejemplo, cómo afecta la órbita de Júpiter (el planeta de mayor masa del sistema solar) al

movimiento de los demás planetas. Era toda una forma nueva de contemplar la mecánica celeste, que a día de hoy es de una idoneidad más que suficiente para pilotar un cohete a la luna.

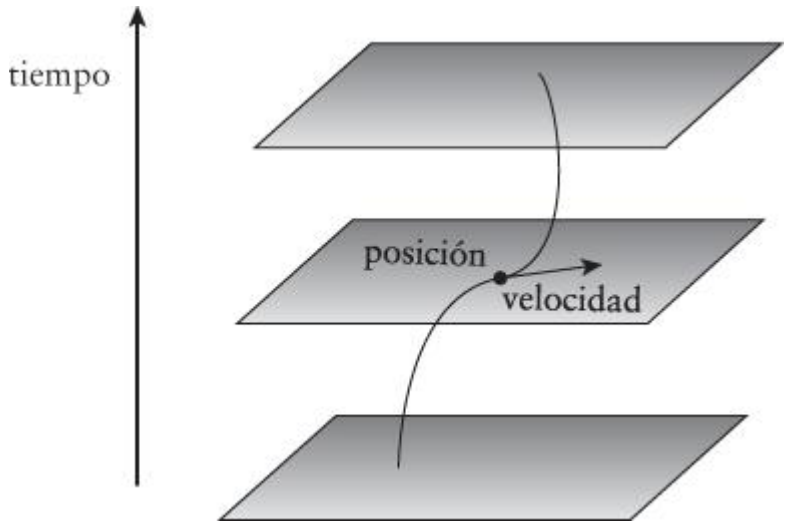
PENSAR EL DESPLAZAMIENTO LOCAL

Nuestro objetivo no es en estos momentos el de repetir cómo Newton derivó de las leyes de Kepler las suyas propias, sino poner de relieve las diferencias filosóficas entre dos modos de abordar la dinámica de los planetas (y, por tanto, la física en general). Kepler nos explica que los planetas se mueven siguiendo una trayectoria elíptica y nos dice algunas cosas acerca de la velocidad a la que se desplazan. Describir una órbita en forma de elipse supone realizar una enunciación global, en cuanto se refiere a la trayectoria considerada como un todo. Esperamos a que el planeta complete una revolución entera alrededor del sol y alcance el punto en que podemos verificar que realmente ha descrito una trayectoria elíptica.

Newton, por su parte, llega a la misma conclusión, pero su procedimiento es muy diferente. El término de partida es ahora local en el tiempo: toda la información relevante se nos presenta en un momento puntual. Primero nos dicen dónde está el planeta, cuál es su velocidad y qué fuerzas actúan sobre él, todo ello en un momento concreto. Luego, la segunda ley de Newton nos informa acerca de la aceleración que el planeta está experimentando en ese mismo momento del tiempo. A partir de ahí, podemos inferir su comportamiento entero en otros momentos puntuales.

Se trata del paradigma laplaciano en acción. Newton lo desarrolló, pero Laplace subrayó su alcance filosófico, y otros investigadores como William Rowan Hamilton aportaron actualizaciones significativas a la idea. El paradigma laplaciano sostiene que toda la información que necesitamos para determinar qué le sucederá al sistema o qué le sucedió en el pasado está contenido en el estado del sistema en cada momento del tiempo. En el caso de un sistema simple como el formado por los planetas y el sol, el estado se halla determinado sencillamente por la posición y la velocidad de

cada uno de sus integrantes. Podría usted objetar que también necesitamos conocer las fuerzas que actúan sobre los planetas, y ciertamente así es; pero tales fuerzas vienen determinadas por las posiciones y velocidades de las demás partes del sistema. De modo que si contamos con la posición y la velocidad, no solo de un planeta, sino de todo lo que hay en el sistema solar, ya podemos seguir adelante.



Para completar este programa, hay dos cosas importantes que tenemos que saber hacer. En primer lugar, hemos hablado de la razón de cambio de varias cantidades. La velocidad es la razón de cambio de la posición; la aceleración es la razón de cambio de la velocidad. ¿Qué queremos decir con esto? ¿Y cómo la calculamos? Si viajamos de un lugar a otro, podemos calcular nuestra velocidad media simplemente dividiendo la distancia total recorrida por el tiempo empleado en el viaje. Pero lo que queremos es la velocidad instantánea, así como la aceleración instantánea, en cada momento de tiempo. Esto no lo obtenemos dividiendo una distancia finita por ningún tiempo finito. Tendremos que ser más listos.

La otra cosa importante es que, suponiendo que conocemos la posición, la velocidad y la aceleración iniciales, queremos ensamblarlas para reconstruir el viaje entero. En otras palabras, queremos determinar la distancia acumulada recorrida a partir de la ratio a la que viajábamos en cada momento. Una vez más, simplificando las cosas, en el caso en que

viajáramos a una velocidad estrictamente constante, no tendríamos más que multiplicar la velocidad por el tiempo de viaje para obtener la distancia recorrida. Pero si a lo largo del trayecto hemos ido acelerando y reduciendo la velocidad, o nos hemos desviado en direcciones diferentes, las cosas ya no resultan tan fáciles.

Estas dos cuestiones —cuál es la razón de cambio instantánea de una cantidad y cómo acumulamos cantidades con el paso del tiempo— son precisamente el objeto del cálculo. En la jerga matemática, se llaman respectivamente derivadas e integrales. Ha llegado el momento de arremangarnos.

FUNCIONES

Supongamos un coche que circula por una carretera recta. Su posición, velocidad y aceleración se representan mediante vectores, pero dado que solo hay una dirección de movimiento y por tanto únicamente un componente de cada vector, podemos darles tratamiento de simples números (que pueden ser positivos o negativos). Imaginemos además que el coche está provisto de un cuentakilómetros de precisión absoluta, que determina exactamente cuál es su posición y que podemos de algún modo registrar esta información para cada instante puntual de tiempo.

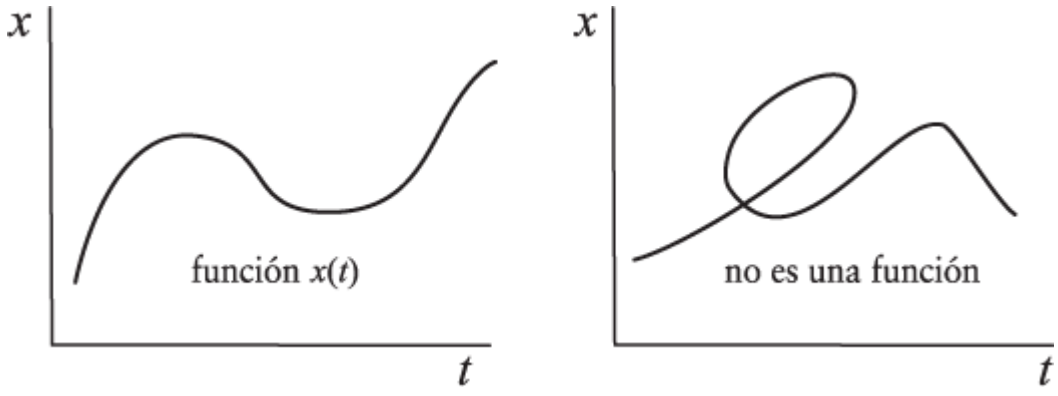
Hablando en términos matemáticos, lo que tenemos es una función: la posición del coche, x, como una función de tiempo, t. Una función es simplemente una correspondencia establecida de una cantidad a otra cantidad: tomo una muestra de la primera cantidad, la traspaso a la función y esta me devuelve un valor de la segunda cantidad. Las cantidades en cuestión pueden ser simples números, series de números o algo más complicado. El número entrante (input) se llama argumento de la función y el obtenido (output) es el valor correspondiente a ese argumento.

Función: argumento → valor

$f: t \rightarrow x$

Una función que asigna t a x podría expresarse como x = f(t), o simplemente x(t) para abreviar. Las letras que utilicemos para señalar las funciones y variables dependen por completo de nosotros, con tal de que recordemos lo que significan. Si denotamos las posiciones de un terreno utilizando las coordenadas x e y, podemos expresar la altura del terreno en cada posición mediante la función h(x, y). De modo que a veces una cantidad como x puede ser el argumento (input) de una función, mientras que otras veces puede desempeñar el papel de valor (output).

En el caso de nuestro coche, el argumento es la variable temporal t, y la función relevante es x(t), el valor de la posición del coche en cada momento de tiempo. Este valor puede repetirse para diferentes argumentos —nuestro coche puede pasar por un punto, dar media vuelta y volver a pasar por él—, pero es mejor que para cada t tengamos un valor específico de x. Esto significa que cuando tracemos una función, la curva podrá ir hacia arriba o hacia abajo, pero nunca dibujará una regresión de izquierda a derecha.

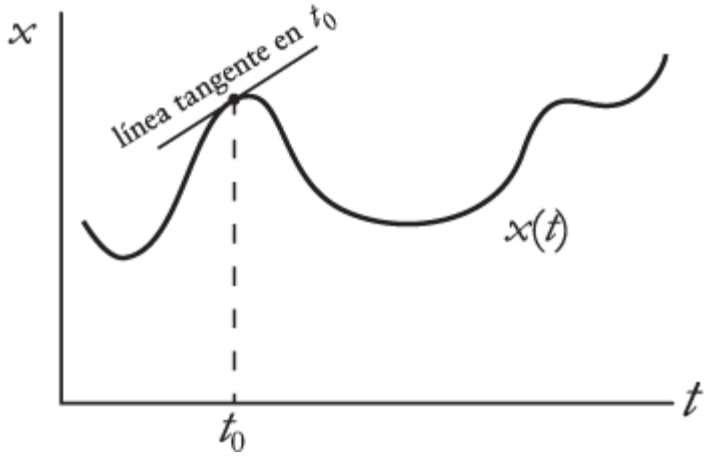


DERIVADAS

Dada x(t), posición como una función de tiempo, podemos preguntarnos cuál es la velocidad en cualquier momento. Sabemos que la velocidad se define como la razón a la que cambia la posición, pero ¿cómo la calculamos a partir de x(t)? La velocidad no puede determinarse solo con saber la posición en un determinado momento: dada la ubicación de un coche sin más información que esa, no tenemos ni idea de cuál es su velocidad. De algún modo tenemos que aprovechar la información acerca de su posición en otros momentos.

Con solo mirar el gráfico de la función, tenemos la impresión general de que la velocidad del coche está relacionada con la pendiente de la curva en cada punto: cómo son de pronunciadas las subidas y bajadas. Donde la curva es prácticamente plana, el tiempo pasa, pero el coche no se mueve mucho: la velocidad es pequeña. Donde la curva es pronunciada, el coche se mueve mucho en poco tiempo: la velocidad es grande.

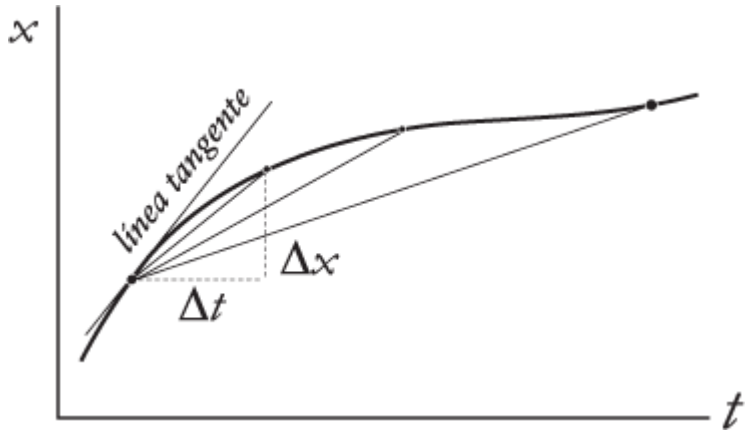
Queremos por tanto imaginar una línea recta que toca exactamente nuestra curva x(t) en algún tiempo t0, llamada línea tangente en ese punto. La velocidad del coche en el momento preciso t0 es la pendiente de esa línea tangente en particular. Necesitamos desarrollar un procedimiento sistemático para definir y calcular la pendiente de la línea tangente en cada punto.



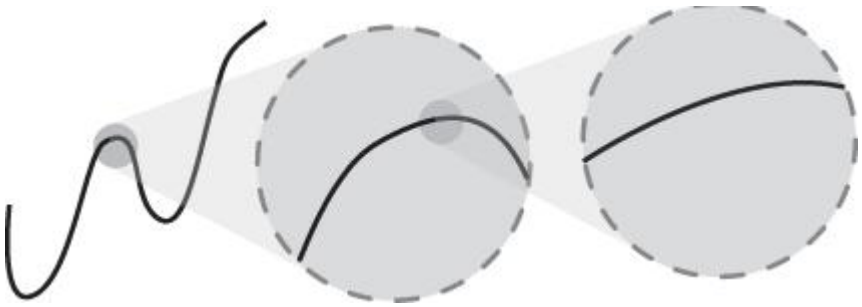
Sería fácil hacerlo si el coche se moviera a velocidad constante. En ese caso la función sería simplemente una línea recta, como en la figura siguiente.

Entonces la pendiente, y por tanto la velocidad, se calculan de un modo sencillo: el cambio de posición dividido por el correspondiente cambio de tiempo. Expresado en símbolos, representemos el cambio de posición como Δx y el cambio de tiempo como Δt, donde Δ es la letra griega delta mayúscula, la cual suele utilizarse para representar el incremento de una cantidad. (La notación Δx representa una cantidad individual simple, el cambio en x, no el producto de cierta cantidad Δ por la cantidad x). Entonces, la velocidad a lo largo de un trayecto en línea recta es:

$$v = \frac{\Delta x}{\Delta t} \qquad (2.3)$$



Llegamos ahora a una de las ideas reveladoras básicas del cálculo. Mientras nuestra función es relativamente uniforme, sin demasiados altibajos ni saltos azarosos de un valor a otro, parecerá recta vista a intervalos muy cortos, aunque tomada en general sea bastante curva. Si vamos amplificándola gradualmente, una línea curva comienza a parecerse cada vez más a una línea recta.



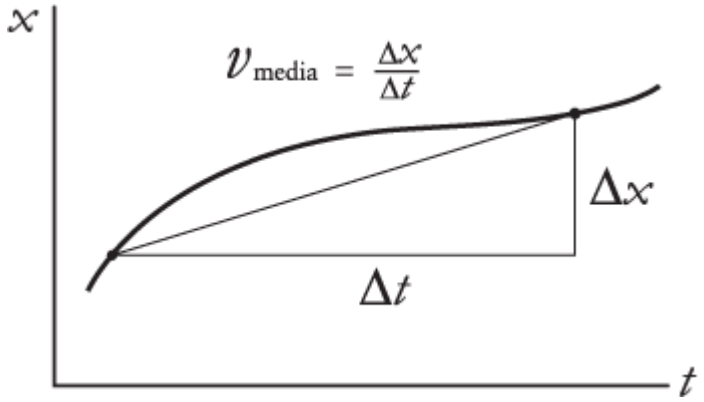
Esto nos sugiere una estrategia. Tomemos un tiempo t, cuya velocidad nos gustaría calcular. Determinemos ahora cierto incremento de tiempo Δt, y consideremos el intervalo de tiempo entre un tiempo inicial t y un tiempo final t + Δt. Dada nuestra función, conocemos tanto la posición x(t) del coche en el tiempo inicial, como su posición x(t + Δt) en el tiempo final. Podemos por tanto calcular el cambio total de la posición a lo largo de todo este tiempo, que es simplemente:

$$\Delta x = x(t+\Delta t) - x(t) \qquad (2.4)$$

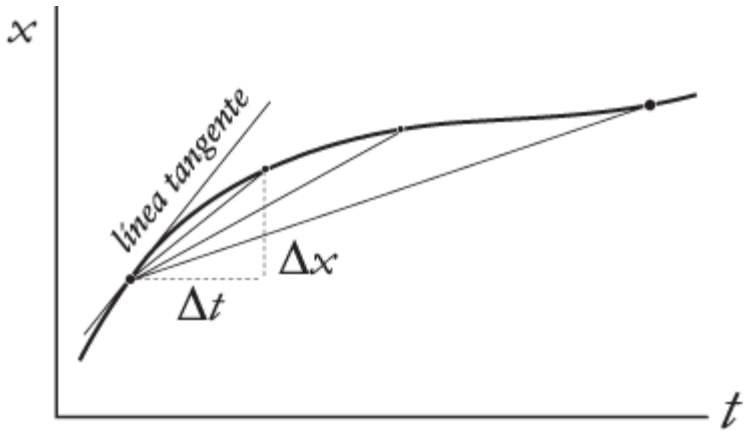
Si la función en su totalidad es curvada más que recta, podemos dividir el cambio total de posición por el cambio total de tiempo, para hallar la velocidad media en ese intervalo.

$$v_{\text{media}} = \frac{\Delta x}{\Delta t} \qquad (2.5)$$

Se parece mucho a (2.3), pero en aquella expresión se refería la velocidad única a lo largo de una trayectoria a velocidad constante, mientras que esta corresponde a la velocidad media para un intervalo de tiempo específico a lo largo de cualquier trayecto dado.



No es esto lo que buscamos. No queremos la velocidad media de un intervalo arbitrario cualquiera, queremos la velocidad instantánea en cada uno de los instantes. Aunque quizás usted ya ve adónde vamos. La elección del intervalo de tiempo Δt ha sido arbitraria, podemos hacer con él lo que queramos. Aumentémoslo. Podemos elegir valores cada vez más pequeños para Δt, lo cual implicará valores cada vez más pequeños para Δx. Mientras que tanto Δx como Δt pueden reducirse hasta llegar al valor cero, su razón Δx/Δt se acerca a una cifra que no tiene que ser cero. De hecho se acerca exactamente a lo que estamos buscando: la pendiente de la línea tangente en nuestro punto inicial.



El procedimiento que acabamos de esbozar se llama determinar el límite cuando Δt tiende a cero. Cero dividido por cero no es nada, matemáticamente hablando; no tiene sentido. Pero podemos determinar el

límite de Δt y Δx cuando ambos tienden individualmente a cero, y su razón —la velocidad, v— está definida. Llamamos a esto la derivada de la función x(t), y lo expresamos con una estimulante notación:

$$v = \text{límite}\left(\frac{\Delta x}{\Delta t}\right) = \frac{dx}{dt} \qquad (2.6)$$

Ya lo tenemos. Esto es una derivada: la pendiente de una curva en un punto, definida por la determinación del límite de la pendiente de una secuencia de líneas que se acercan cada vez más a la línea tangente de ese punto. Considerábamos el caso particular de x como función de t, en que la derivada es la velocidad, pero la idea es mucho más general. La aceleración, por ejemplo, es la derivada de velocidad con respecto al tiempo:

$$a = \frac{dv}{dt} \qquad (2.7)$$

La velocidad se mide en metros por segundo, mientras que la aceleración se mide en metros por segundo al cuadrado, por cuanto es la razón de cambio de la velocidad. La aceleración debida a la gravedad de un objeto en caída libre cerca de la Tierra es de 9,8 metros por segundo al cuadrado (abreviada a veces como m/s^2).

Con frecuencia nos interesa alguna función de x, que podemos expresar como f(x), y su derivada será df/dx. Toda función lo es «de» alguna variable, y podemos calcular la derivada respecto de esa variable. No importa en absoluto qué símbolo utilicemos para referirnos a la variable, es tan solo una etiqueta que elegimos por conveniencia. Nos resulta útil para una más rápida comprensión llamar al tiempo t y a la distancia x, pero depende enteramente de nosotros.

Las cantidades dx y dt se conocen como infinitesimales. Parecen números que dividimos para obtener v, pero la realidad es un poco más sutil. Si

fueran realmente números, su valor sería cero, y eso no sirve. Representan más bien la idea del límite de Δx y Δt cuando estas dos cantidades se aproximan a cero. La razón de estos dos infinitesimales es un número bien definido, aunque individualmente no lo sean. Todo lo que puedo decir es que los matemáticos han desplegado una enorme cantidad de pensamiento para hacer que todo esto suene respetable. Como físicos, tendemos a no preocuparnos demasiado. Si funciona, pulgares arriba y pasemos al problema siguiente.

Hay dos cosas que pueden estar incomodándolo llegados a este punto. En primer lugar, parece casi demasiado fácil. Todo lo que hemos hecho ha sido definir la derivada poniéndonos quisquillosos en torno a la pendiente de una línea tangente a una curva. Algo con una reputación tan intimidadora como el cálculo debería ser más difícil que eso. Y en segundo lugar, cuando se acaba la historia sigue sin estar del todo claro qué se supone que vamos a hacer con esa información. Todo ello es un poco abstracto. Si nos dan una función real, o la lectura de un cuentakilómetros, ¿de verdad tenemos que pasar por todo este embrollo para hallar su derivada?

Ambas cuestiones están relacionadas, y básicamente se anulan la una a la otra. Si estuviéramos en una clase de cálculo real, se vería usted inmerso en un tedioso pero explícito y tranquilizador conjunto de reglas acerca de cómo calcular realmente las derivadas de unas funciones dadas, proceso conocido como diferenciación. Consideremos una función muy simple, $f(x) = ax + b$, donde tanto a como b representan parámetros fijos (a menudo llamados simplemente constantes).[2] Esto se conoce como una función lineal, puesto que al trazarla en un plano daría una línea recta.

Podemos hallar la derivada de esta función lineal si lo pensamos un poco. La constante b no afecta a la pendiente de la línea, de modo que podemos ignorarla. Y la constante a simplemente es la pendiente: si cambiamos x por una cantidad Δx, entonces $f(x)$ cambia a $a\Delta x$, y así $\Delta f(x)/\Delta x = a$, sin importar lo que sea x. Tenemos por tanto:

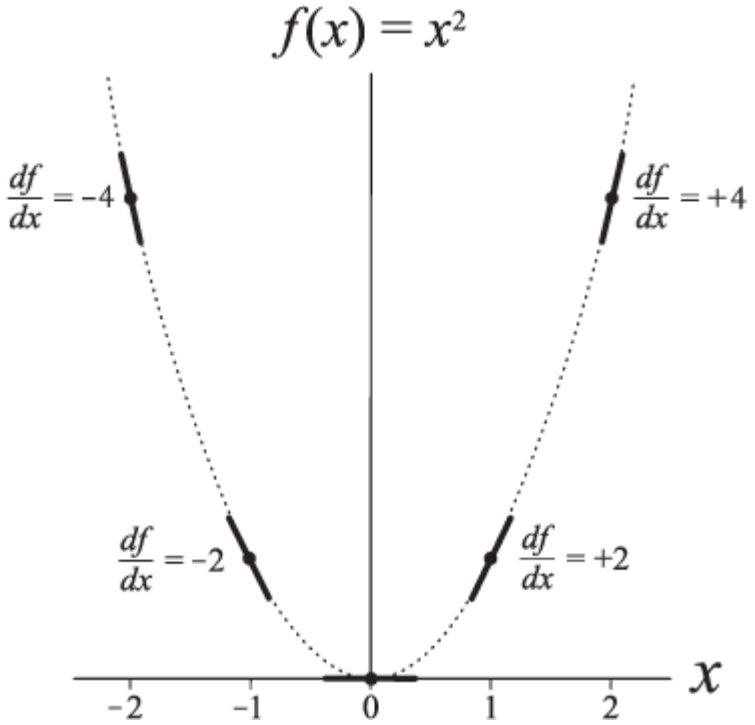
$$\frac{d}{dx}(ax+b)=a \qquad (2.8)$$

La derivada de nuestra función lineal es simplemente una constante, igual que el número que multiplicaba x en la función original.

Pero la diferenciación no suele ser tan simple. La mayor parte de las curvas tienen una pendiente que cambia a cada punto. La derivada de la función parabólica $f(x) = x^2$, por ejemplo, viene dada por:

$$\frac{d}{dx}x^2 = 2x \qquad (2.9)$$

Podemos ver esto en la figura, que representa en un plano la función $f(x) = x^2$, así como la pendiente de las líneas tangentes a ciertos puntos ($x = -2, -1, 0, 1, 2$). Para valores negativos de x, la parábola desciende en pendiente y su derivada es negativa. Para valores positivos de x, la derivada se hace cada vez más positiva a medida que se incrementa.



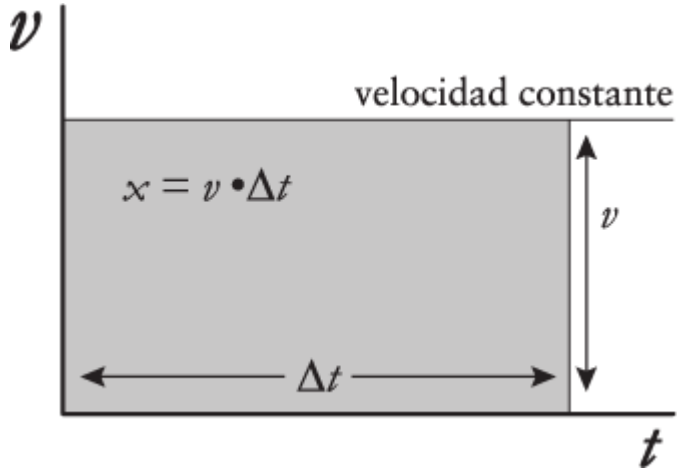
Existen fórmulas análogas para las derivadas de x elevada a otras potencias, y para la raíz cuadrada de x, y el logaritmo de x, y el seno y el coseno de x, y el producto de cualesquiera de estas dos funciones, etcétera. Pueden verse ejemplos en el Apéndice A.

Aprender todas estas técnicas es lo que hace de la clase de cálculo una tarea bastante ardua a veces, pero es también lo que la convierte en algo muy útil para el trabajo del científico. Nuestro objetivo aquí es el de entender el mundo lo mejor que podamos, no el de adiestrarnos para llegar a ser físicos profesionales. Básicamente, tenemos que concentrarnos en las partes divertidas. La diferenciación nos da una estrategia para calcular la pendiente de una curva, por ejemplo, para deducir la velocidad de un coche a partir de su posición en un periodo de tiempo. Con este conocimiento en mano, podemos seguir adelante.

INTEGRALES

De acuerdo con el paradigma laplaciano de la física clásica, podemos determinar la evolución futura de un objeto si conocemos su posición presente y su velocidad, así como estas mismas cantidades para todos los demás objetos relevantes que puedan estar afectándolo. Podemos determinar por tanto qué fuerzas están actuando sobre el objeto y, a partir de la segunda ley de Newton (2.1), podemos hallar su aceleración. Provistos de todo esto, nos gustaría determinar realmente la trayectoria del objeto. Si la velocidad es la razón de cambio de la posición, y la aceleración es la razón de cambio de la velocidad, nos gustaría sumar todos estos cambios acumulados para determinar cómo evolucionan la posición y la velocidad en el transcurso del tiempo.

Volvamos al ejemplo del coche que transita en línea recta y a velocidad constante. Pero en lugar de plantear la posición como una función de tiempo, supongamos no saber esto y planteemos la velocidad como una función de tiempo. Es muy fácil cuando la velocidad es constante.

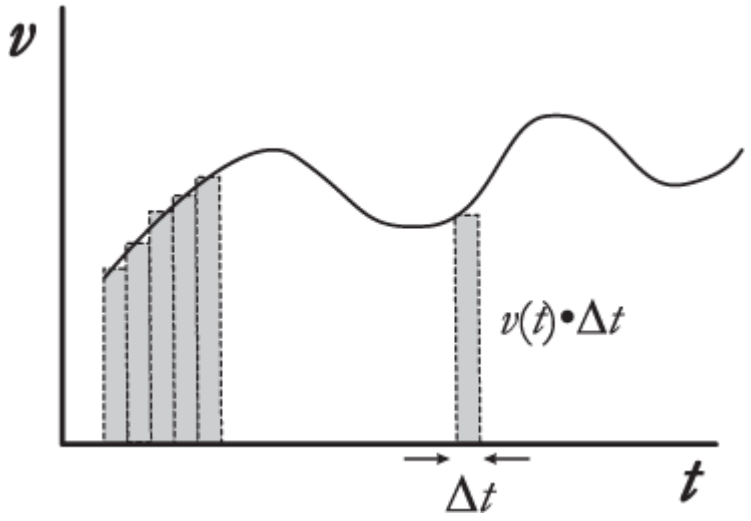


Cuando algo se desplaza a una velocidad constante v, la distancia x recorrida resulta del producto de esa velocidad por el tiempo que tarda en desplazarse: x = v. Hay una bonita forma geométrica de pensar esto en los términos representados en al diagrama anterior: la distancia recorrida es simplemente el área del rectángulo que va en horizontal de 0 a Δt y en vertical de 0 a v. Podemos considerar la suma acumulada de una cantidad como el área representada bajo la curva definida por su función.

Nos gustaría poder generalizar esto para el caso en que la velocidad no es constante. Cuando estábamos desarrollando la idea de una derivada, el truco era amplificar una curva hasta que pareciera una línea recta. Podemos recurrir aquí a un razonamiento análogo. Podemos aproximarnos al valor del área limitada bajo una función curvilínea v(t) tomando un pequeño incremento Δt, calculando el área formada por el rectángulo de t = 0 a t = Δt (que es simplemente v(0)·Δt), y luego haciendo lo mismo para el tramo en la curva que va de t = Δt a t = 2Δt, y así sucesivamente. Al final sumamos todas las áreas de estos finos rectángulos y obtenemos una cantidad aproximadamente igual al área limitada bajo la curva. El procedimiento de sumar una sucesión de cantidades se representa mediante el símbolo sumatorio ∑, la letra griega sigma mayúscula.

$$\text{área bajo } v(t) \approx \sum_{\text{rectángulos}} v(t)\Delta t \qquad (2.10)$$

El signo de igual formado por dos líneas ondulantes es una forma de expresar «aproximadamente igual a».



A medida que nuestro intervalo de tiempo Δt va haciéndose cada vez más pequeño y nuestros rectángulos cada vez más finos (pero también más numerosos), el proceso se aproxima más y más al área real limitada bajo la curva. Así, como en el caso de las derivadas, tomamos el límite de Δt según se hace infinitesimalmente pequeño y lo llamamos dt. Remplazamos también el signo sumatorio ∑ por un nuevo símbolo con el aspecto de una S artísticamente estilizada. El resultado es la integral de velocidad con respecto al tiempo, que nos da la distancia acumulada recorrida:

$$\Delta x(t) = \text{límite}\left[\sum v(t)\Delta t\right] = \int v(t)\,dt \qquad (2.11)$$

De forma similar, la cantidad con que cambia la velocidad puede calcularse como la integral de la aceleración:

$$\Delta v(t) = \int a(t)dt \qquad (2.12)$$

Estamos ocultando algo de información en aras de la claridad. Lo que hemos calculado es, en cambio, en la posición (o velocidad) en el intervalo de cierto tiempo inicial y cierto tiempo final; una notación más cuidadosa indicaría cuáles eran esos dos puntos temporales. Véase el Apéndice A para mayor detalle.

Al igual que podíamos pensar las derivadas como una forma de dar sentido a cero dividido entre cero, podemos pensar una integral como una forma de dar sentido a infinito multiplicado por cero, donde infinito es el número de finos rectángulos inscritos bajo la curva y cero es el área de cada uno de los rectángulos. Podemos buscar un resultado francamente riguroso (los matemáticos se refieren a este tema simplemente como análisis). Pero puede intuir por qué Newton dudaba en confiar en el cálculo en los Principia. Estas ideas eran muy nuevas en su época, e incluso hoy algunas personas que trabajan en los fundamentos de las matemáticas se muestran escépticas acerca de que el cálculo sea totalmente riguroso. Por fortuna funciona más que suficientemente bien para los propósitos de la física.

Contando con las leyes de Newton y conociendo las fuerzas que actúan sobre el objeto, no necesitamos nada más para obtener la aceleración: nos la da directamente F = ma. Podemos agregar la aceleración para obtener la velocidad, y podemos agregar la velocidad para obtener la posición. Vemos así cumplida la promesa del paradigma laplaciano: dadas una posición y una velocidad iniciales, podemos reconstruir la trayectoria completa del objeto.

Podemos pensar en la diferenciación y la integración como operadores que actúan sobre funciones: aplicaciones de unas funciones a otras funciones. Si me dan una función, puedo crear otra tomando su derivada, y otra más tomando su integral. De hecho, son operaciones inversas una con respecto de la otra; la derivada contrarresta la integral, y viceversa.

$$\text{derivada}\{\text{integral}[f(x)]\} = f(x)$$
$$\text{integral}\{\text{derivada}[f(x)]\} = f(x) \qquad (2.13)$$

O expresado en símbolos:

$$\frac{d}{dx}\int f(x)dx = f(x)$$
$$\int \frac{df}{dx}dx = f(x) \qquad (2.14)$$

En la práctica, calcular derivadas de funciones es bastante fácil, pero inferir sus integrales puede ser bastante difícil. Hasta los mejores físicos recurren a menudo al ordenador para hallar la solución numérica cuando necesitan saber el valor de una integral en particular.

No nos preocupa demasiado «hacer» derivadas e integrales (tomar una función específica y calcular la derivada o la integral). Se facilitan algunos ejemplos en el Apéndice A, pero lo que nos interesa principalmente son los conceptos implícitos, más que los cálculos explícitos. Utilizaremos un simple resultado: la integral del intervalo infinitesimal es igual a un intervalo finito:

$$\int dx = \Delta x \qquad (2.15)$$

Esta ecuación simplemente dice: «La cantidad total de x que acumulamos es el cambio en x desde el principio hasta el final del proceso». Aquí, la variable x puede valer para cualquier cosa: la distancia en el espacio, la duración en el tiempo, cualquier magnitud física que nos interese en tal momento.

CONTINUIDAD E INFINITO

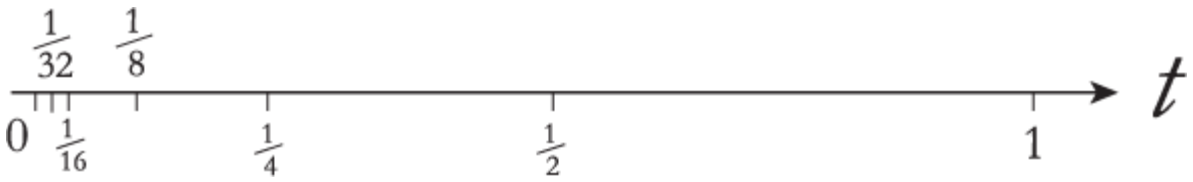
Desde el día en que Isaac Newton estableció sus leyes, se han propuesto numerosas otras leyes relacionadas con sistemas físicos fundamentales. James Clerk Maxwell dispuso una gama de ecuaciones en torno a la electricidad y el magnetismo; Albert Einstein propuso una ecuación para la curvatura del espacio-tiempo; Erwin Schrödinger sugirió una ecuación para la función de onda de un sistema cuántico-mecánico, y muchas más. Lo que todas ellas tienen en común es que son ecuaciones diferenciales, incluyen derivadas (con respecto al tiempo, y con frecuencia también con respecto al espacio) de cualquier sustancia objeto de descripción. Por eso el cálculo es tan central para la manera en que opera la física.

¿Podría ser de otro modo? No conocemos las leyes finales de la física, por lo que deberíamos estar abiertos a diferentes posibilidades mientras seguimos pensando en cuáles podrían ser. Una posibilidad, por supuesto, es que el paradigma laplaciano entero carezca de base, que las leyes fundamentales sean intrínsecamente globales en lugar de locales, de forma que no pudieran pensarse siguiendo el esquema de partir de una información en un momento concreto y avanzar o retroceder afanosamente para elaborar la solución entera.

Otra posibilidad sería que el tiempo fuera discreto, en lugar de continuo. Quizás exista realmente un intervalo de tiempo máximamente breve y el universo evolucione en unidades de ese intervalo, en lugar de hacerlo sin fisuras. Es una posibilidad a contemplar bastante obvia, y que cuenta con ventajas y desventajas. Una ventaja estriba en el hecho de haber numerosos enigmas asociados a la idea de continuidad (y a la de su prima hermana, la idea de infinito); quizá el motivo de que resulten tan enigmáticos sea que no se adecúan al mundo real: una preocupación menos, entonces. Una desventaja es que tendríamos que arrojar por la borda todo cuanto pensábamos acerca de la mecánica clásica y de la relatividad, al menos como constitutivos de la descripción fundamental del mundo. En último

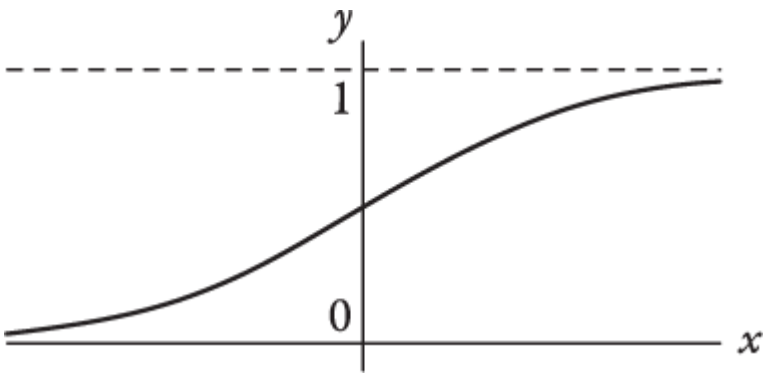
término valdría la pena hacerlo, pero lo razonable es abordar tan radicales cambios con cierto grado de cautela.

Decir que el tiempo (o cualquier otra cantidad) es continuo implica que contiene un número infinito de valores, incluso entre dos puntos separados por una distancia finita. Podrá convencerse si imagina una línea que represente el tiempo que va del pasado al futuro. Seleccione dos instantes, que puede señalar como t = 0 y t = 1.



Es evidente que podemos marcar un instante en el punto medio entre ambos, señalándolo como t = ½. Pero entonces es igualmente evidente que podemos marcar otro instante en el punto medio entre t = 0 y t = ½', que será t = ¼. Y podemos proseguir así con t = ⅛, 1/16, 1/32… Y no teníamos por qué tomar t = 0 y el número siguiente, sino que podíamos haber tomado t = 1 y el número anterior, obteniendo así t = ¾, ⅞, 15/16, 31/32 … Fácilmente podemos ver que hay un número infinito de números entre dos puntos cualesquiera de una línea continua.

Lo que resulta más destacable es que el infinito que representa «el número de números entre 0 y 1» es del mismo tamaño que el infinito que representa «el número de números entre −∞ y +∞». Esto parece extraño, porque tendemos a pensar que un subconjunto de algo debería tener siempre menos elementos que el conjunto entero. Pero el infinito es especial. Sabemos que hay un número igual de números entre 0 y 1 que entre −∞ y +∞ porque podemos establecer, uno a uno, una correspondencia exacta entre ellos. Esta correspondencia puede representarse mediante una simple función:



Tenemos una función de x a y, con la propiedad de que a cualquier valor de x entre $-\infty$ y $+\infty$ se le asigna un valor de y entre 0 y 1.

Todo esto puede llevarle a pensar que todas las cantidades infinitas son en definitiva del mismo tamaño. Ciertamente, si multiplica infinito por 2 (o por cualquier otro número mayor de 0), obtiene el mismo infinito. El número de enteros es igual al número de enteros pares. Pero las cosas no son tan sencillas: hay diferentes grados de infinitud, como descubriera el matemático alemán Georg Cantor a finales del siglo XIX. El teorema de Cantor muestra que hay infinitos números enteros e infinitos números reales, pero que el número de números reales es mayor que el de números enteros. Este descubrimiento no fue acogido con universal entusiasmo. Muchos matemáticos dudaron del resultado de Cantor, y su contemporáneo Leopold Kronecker lo tildó de «corruptor de la juventud». Los matemáticos modernos aceptan en general la validez del argumento de Cantor, pero siguen indecisos a la hora de decir si deberíamos confiar en las conclusiones que implica.

¿Es todo esto relevante para la física? Tal vez no. Los seres humanos somos criaturas finitas, y una persona puede adoptar la postura de pensar que ninguno de nosotros puede realmente establecer una diferencia práctica entre «infinito» y «enormemente grande» (o entre «cero» y «enormemente pequeño»). De modo que, cuando se trata de describir el mundo físico, quizá podamos elegir entre ocuparnos o no del infinito. Pero no deberíamos confundir lo que los seres humanos podamos tener en la mente con lo que hay realmente en la naturaleza. Es concebible que algún día apreciemos

que, en lo tocante a la realidad, hay una manera mejor de pensar sobre las cuestiones de la continuidad y el infinito, aunque aún no la conozcamos.

1 Rocky Kolb, Blind Watchers of the Sky: The People and Ideas That Shaped Our View of the Universe , Basic Books, 1997, p. 134.

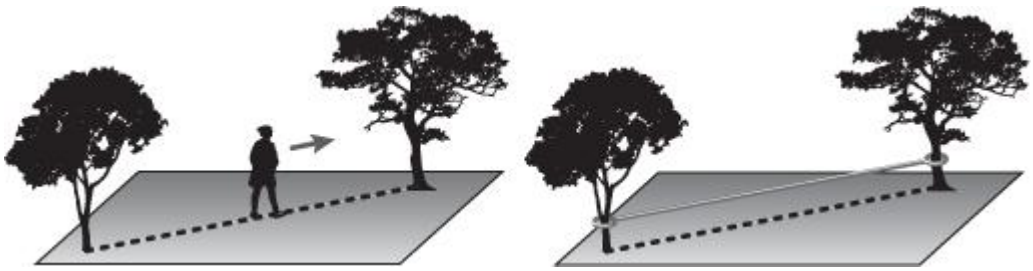
2 Aquí x es el argumento de la función, y f (x) es su valor. No hemos asignado un nombre variable separado para el valor, pero podría hacerlo si quisiera, por ejemplo, y = f (x). Si tuviéramos que trazar la función en un plano, x sería el eje horizontal y f (x) sería el eje vertical.

TRES

DINÁMICA

Imagine dos árboles, separados a cierta distancia uno de otro, situados en un parque de suelo perfectamente plano. De pie y con la espalda apoyada en uno de los árboles, oriéntese hacia la dirección del otro árbol. Póngase ahora una venda en los ojos y camine. Supongamos que es capaz de mantener la dirección en que se está desplazando y que nadie le pone trampas ni se interfiere en su camino. Al final del trayecto, se encontrará justo debajo del otro árbol. Y si se quita la venda de los ojos y las huellas que ha dejado, comprobará que se ha movido en línea recta.

Va a hacer ahora algo completamente diferente. Coja una cuerda larga, ate un extremo a uno de los árboles y lleve el otro extremo hasta el otro árbol. Ate la cuerda bien tensa al tronco, de modo que hayamos utilizado la menor cantidad de cuerda posible para unir ambos árboles. El resultado es nuevamente una línea recta: la cuerda estirada habrá quedado exactamente encima de la línea formada por sus huellas durante el ejercicio anterior.



Todo esto resulta bastante obvio, y al mismo tiempo es notable de algún modo. Tenemos una noción de «línea recta» basada en el sentido común, pero dos formas distintas de construirla. Una de ellas consistiría en «desplazarnos en todo momento en la misma dirección», y la otra en

«minimizar la distancia entre el punto inicial y el punto inicial». La primera forma es reflejo de una filosofía local de la acción, en el espíritu del paradigma laplaciano del que hablamos en el capítulo anterior: en cada momento de tiempo estoy haciendo una cosa en particular, y al final todo mi trabajo ha formado determinado camino. La segunda forma es más global, reminiscencia de las leyes de Kepler: de todas las maneras posibles en que podía tenderse una cuerda entre dos árboles, he elegido la más corta. Solo que estas dos ideas, aparentemente independientes, ofrecen la misma respuesta a fin de cuentas.

Los físicos trabajan del mismo modo. Hemos destacado, por buenos motivos, la idea de Laplace de que la información contenida en un sistema en determinado estado y en un momento concreto puede utilizarse para reconstruir su evolución entera en el tiempo, avanzando implacablemente paso a paso de un momento al siguiente. Pero hay otras formas de obtener la misma respuesta, en última instancia equivalente, aunque muy diferente en apariencia. Estas otras formas pueden apelar en su formulación a un conjunto de conceptos fundamentales bastante diferente. Esto suscita preguntas interesantes acerca de cuál sea el vocabulario «mejor» o «más real». Si hay diferentes modos de pensar que son auténtica y rigurosamente equivalentes, quizá no importe demasiado. Pero no conocemos las leyes últimas de la física, y es posible que una forma de pensamiento nos proporcione una vía más directa de llegar hasta ahí. Como dijo Richard Feynman en cierta ocasión, es posible que dos formulaciones de una teoría ofrezcan exactamente las mismas predicciones, pero que no sean «psicológicamente idénticas cuando intenten ir desde esa base hacia lo desconocido».

El capítulo anterior estaba centrado en el «cambio», mientras que este versa acerca de la dinámica. La diferencia radica en que el cambio es un concepto completamente general, mientras que la dinámica se ocupa específicamente de los cambios que obedecen a las ecuaciones de la física. Consideraremos las propiedades de sistemas físicos específicos y que nos dice la mecánica clásica acerca de su comportamiento. La reflexión en torno a la energía cinética y la energía potencial nos conducirá a nociones interesantes que relacionan la dinámica de objetos diferentes. Ello nos estimularé en último término a reformular la mecánica a través de una nueva visión global que

considera la historia del sistema como un todo, una relevante idea conocida como el «principio de mínima acción».

LO QUE IMPORTA PARA LA ACCIÓN

Pensemos un poco más sistemáticamente en lo que sucede en el paradigma laplaciano. Para simplificar las cosas, nos centraremos en una partícula individual moviéndose en un espacio tridimensional, y cuya posición está especificada por un vector . El estado del sistema consiste en esa posición juntamente con la velocidad de la partícula, que es la derivada de posición con respecto al tiempo, . Seis números en total, para una partícula en tres dimensiones: tres componentes de posición más tres componentes de velocidad.

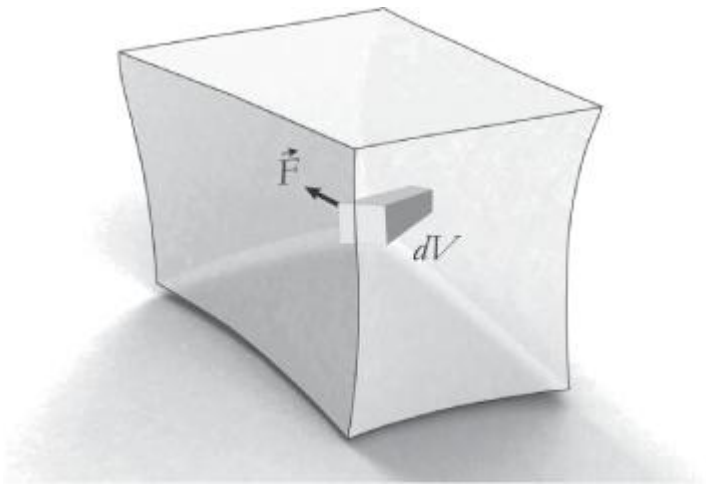
Así que la receta es esta. Tenemos cierta situación específica, como «una pelota rodando por una colina» o «un planeta en órbita alrededor del sol». Me dan los datos de un momento concreto de tiempo t0. A partir de ahí, y de nuestro conocimiento de la situación, podemos calcular la fuerza total que actúa sobre el objeto: . Si nuestro objeto representa un planeta, por ejemplo, obtenemos la fuerza aplicando la ley de gravitación de Newton al sol y a todos los demás planetas. Luego, la segunda ley de Newton nos proporciona la aceleración: . Ahora sabemos, no solo los datos iniciales , sino también con qué rapidez cambian:

$$\text{velocidad } \vec{v} = \text{razón de cambio de } \vec{x} = \frac{d\vec{x}}{dt}$$

$$\text{aceleracción } \vec{a} = \text{razón de cambio de } \vec{v} = \frac{d\vec{v}}{dt} = \frac{\vec{F}}{m}$$

A partir de ahí podemos utilizar el cálculo para avanzar en el tiempo, reconstruyendo la trayectoria entera [(t), (t)].

Es una estructura maravillosamente flexible. Hemos estado hablando mucho de partículas, pero la mecánica clásica es mucho más general. Pongamos que queremos describir un material extenso, como un sólido, un líquido o un gas. E imagine que estamos considerándolo macroscópicamente, como materia verdaderamente continua en lugar de como una colección de átomos individuales. Lo que podemos hacer es concentrarnos en una diminuta parte del objeto y amplificarla, para obtener un «elemento de volumen» infinitesimal de material: dV. Sobre este elemento actuarán diversas fuerzas; experimentará la atracción de las fuerzas gravitatoria y eléctrica procedentes del mundo exterior, pero también recibirá el empuje generado por la presión de otra parte del material que lo rodea. Si conocemos la posición y la velocidad de nuestro pequeño elemento, así como las fuerzas que actúan sobre él, las leyes de Newton nos dirán cómo va a moverse. Finalmente, es el cálculo el que viene al rescate: podemos derivar ecuaciones para el sistema considerado como un todo sumando lo que sucede a cada diminuto elemento de volumen.



El hecho de que necesitemos posición y velocidad como datos entrantes para determinar la trayectoria de nuestro sistema es importante, pero también lo es que no necesitemos ninguna otra información sobre el movimiento inicial del sistema. Necesitamos saber cuál es la aceleración, pero no necesitamos que se nos dé como elemento independiente de información, ya que viene determinada por la segunda ley de Newton (una

vez conocemos nuestra situación y qué está haciendo el resto del mundo). La velocidad es la derivada de posición y la aceleración es la derivada de velocidad, por eso decimos que la aceleración es la segunda derivada de posición:

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d}{dt}\left(\frac{d\vec{x}}{dt}\right) = \frac{d^2\vec{x}}{dt^2} \qquad (3.1)$$

Recuerde al leer este tipo de notación que d no es una variable; d/dt es un operador que toma la derivada con respecto al tiempo. En la segunda derivada utilizamos dos veces el operador d/dt sobre la misma función, lo cual escribimos d^2/dt^2.

Podemos también contemplar derivadas de orden superior, a las que se han dado nombres un tanto caprichosos, en inglés:

•Velocidad = primera derivada de posición (con respecto al tiempo)

•Aceleración = derivada de velocidad = segunda derivada de posición

•Jerk («tirón», o sobreaceleración) = derivada de la aceleración = tercera derivada de posición

•Snap («chasquido») = derivada de jerk = cuarta derivada de posición

•Crackle («crujido») = derivada de snap = quinta derivada de posición

•Pop (onomatopeya) = derivada de crackle = sexta derivada de posición

Estos términos son útiles a veces en aplicaciones de ingeniería (y desayunos).[1] Pero los físicos raramente los usan. En la mayoría de los casos simplemente no se necesitan: una vez conocemos la posición y velocidad de cada elemento del sistema, podemos determinar las aceleraciones y seguir a partir de ahí.

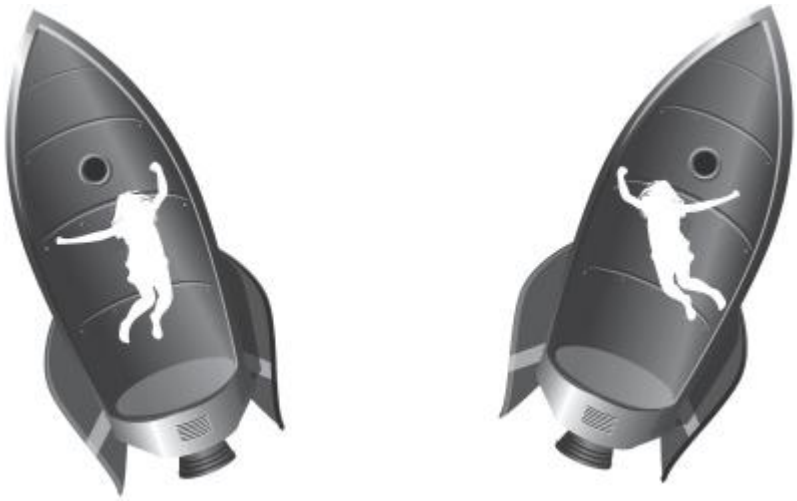
RELATIVIDAD GALILEANA

Hay algo bastante profundo implícito en todo esto. El hecho de que necesitemos que nos den la posición y velocidad de nuestro sistema refleja el hecho de que no existe ninguna posición o velocidad privilegiada en el universo —no hay un estado categorizado como especial, con respecto al cual pudieran medirse otros estados—, al menos en lo que respecta a las leyes de la física. En cuanto a la posición, no es demasiado difícil de aceptar. Si realizamos un experimento de física para someter a prueba una ley fundamental de la naturaleza en Florencia, Italia, y luego llevamos a cabo el mismo experimento en Cambridge, Inglaterra, esperaremos obtener el mismo resultado. Podemos obtener resultados diferentes si medimos cantidades que dependen de nuestra localización, como la velocidad del sonido (que depende de la presión atmosférica) o incluso la aceleración debida a la gravedad (que depende de la altura). Pero no cabe esperar que las leyes implicas en sí mismas, como la segunda ley de Newton o la ley de la inversa del cuadrado para la gravedad, cambien de un lugar a otro. Las leyes de la física no consideran especial ningún lugar en el universo.

Las leyes de la física tampoco distinguen ninguna velocidad como especial, aunque esto resulta menos intuitivo. Cuando mencionamos la velocidad de algo, estrictamente hablando, se mide siempre en relación con otra cosa.

Dados dos objetos, la distancia entre ambos está bien definida, y su velocidad mutua es la derivada de esa distancia con respecto al tiempo. No existe nada a lo que podemos llamar «la velocidad», punto. En nuestra vida diaria, esta profunda característica de la realidad permanece ensombrecida para nosotros, porque nunca nos alejamos demasiado de un marco referencial obvio: la Tierra. Cuando hablamos de la velocidad de un coche o de un avión, asumimos por defecto que dicha velocidad se mide con respecto a la Tierra. Pero eso es una característica de nuestro entorno habitual, no algo intrínseco a las leyes fundamentales de la física. (Y, tal como los pilotos aprenden pronto, es importante distinguir entre nuestra velocidad relativa al suelo y nuestra velocidad relativa al aire).

Imagine que está usted encerrado en el interior de una nave espacial con los motores apagados, sin experimentar aceleración alguna. Sin mirar al exterior (ni usar instrumentos para hacerlo), no tendría forma de saber con qué rapidez se está moviendo. Eso es porque no existe tal cosa: «con qué rapidez se está moviendo», sin especificar con respecto a qué. No hay una medida absoluta de reposo en el universo, como tampoco había un lugar privilegiado.



La ausencia de un valor de reposo privilegiado en las leyes de la física la señaló por primera vez Galileo. Él no estaba pensando en naves espaciales, pero sí que propuso un escenario equivalente recurriendo a los navíos en el

mar. Su planteamiento se integraba en el contexto de la discusión de actualidad en su época acerca de si el sol se movía alrededor de la Tierra, o si era la Tierra (como pensaba Galileo) la que rotaba. Muchas personas argüían que, si la Tierra rotara lo sabríamos, porque el movimiento de la superficie terrestre se sumaría al movimiento de cualquier otra cosa. Galileo replicaba que lo único que contaba era el movimiento relativo entre dos objetos. Postulaba, por ejemplo, que si dejamos caer una bala de cañón desde lo alto del mástil de un barco, caería en línea recta desde el punto de vista de alguien que estuviera en el barco, estuviera el barco quieto o en movimiento con respecto al mar.

La referencia a movimiento «relativo» nos hace pensar en la teoría de la relatividad, y no sin razón. La ausencia de un lugar o de un valor de reposo privilegiados en el universo se conoce como principio de relatividad, y de él habló Galileo mucho antes de que Einstein entrara en escena. El edificio de la mecánica newtoniana descansa sobre los fundamentos de la relatividad galileana, que considera relativa cualquier velocidad entre dos objetos. La moderna teoría de la relatividad sustituye esta noción por la relatividad «lorentziana», del físico holandés Hendrik Antoon Lorentz. La diferencia estriba en que la relatividad lorentziana incluye un límite superior en la velocidad relativa entre dos objetos, determinado por la velocidad de la luz.

Advierta que hemos negado la existencia de una posición privilegiada o de una velocidad privilegiada, pero no hemos negado la existencia de una aceleración privilegiada. El motivo es que sí hay una aceleración privilegiada: cero. Hay un tipo especial de trayectorias, conocidas como trayectorias inerciales, que son aquellas que no están experimentando ninguna aceleración. A diferencia de la posición o la velocidad, si usted estuviera encerrado en una nave espacial, podría decir cuándo esta recibía una aceleración: cuando la recibiera, la nave espacial le empujaría en esa dirección.

LA PELOTA SOBRE LA COLINA

Siguiendo la estela de la filosofía de la vaca esférica, una de las cosas más útiles que podemos hacer en física es determinar un problema ideal simplificado que esté relacionado con un gran número de situaciones más realistas, para entonces comprender ese «modelo de juguete» (como a los físicos les gusta llamarlos) lo mejor que podamos. Todos tenemos un conocimiento intuitivo acerca de cómo funciona el mundo, pero ese conocimiento intuitivo puede desarrollarse y ampliarse familiarizándonos con ejemplos comprensibles.

No hay un modelo de juguete más común y útil en física que una pelota rodando arriba y abajo por un paisaje de colinas. Ya recurrimos a él en el capítulo 1, cuando hablamos de la energía cinética y potencial. Este ejemplo va a resultar mucho más útil de lo que habría podido pensar en un principio. Lo que nuestra percepción aprenda pensando en pelotas que ruedan pendiente abajo será directamente aplicable a los campos cuánticos y al modelo estándar de la física de partículas.

Lo que en realidad tenemos en mente es un modelo de juguete idealizado, a pesar del hecho de que probablemente hayamos entrado en contacto con

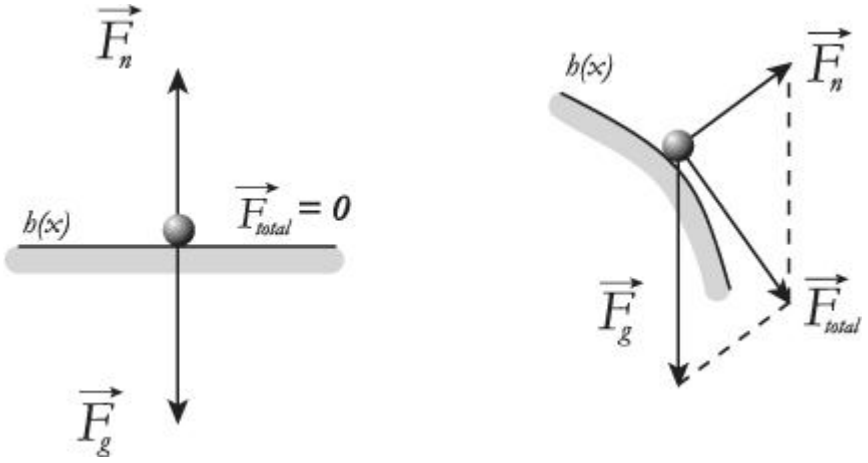
pelotas reales rondando por pendientes de colinas reales en algún momento de nuestra vida. Para la ocasión, como es habitual, ignoraremos completamente la resistencia del aire y la fricción, y demás formas en que la energía puede disiparse. No sería muy difícil incluirlas, pero en nuestro punto de partida intentamos siempre simplificar las cosas hasta lo más esencial, para ir añadiendo complicaciones más tarde. Analizando las cosas con mayor sutileza aún, una pelota real tendría al rodar no solo energía cinética y energía potencial, sino también la energía proveniente de la rotación producida en la acción misma de rodar. Ignoraremos esto también. Lo que imaginaremos es una partícula sin más, moviéndose sin pérdida de energía por un paisaje de energía potencial, y conservando perfectamente en todo momento la suma de energía potencial y energía cinética.

De acuerdo con las reglas de la mecánica clásica, queremos especificar la posición y la velocidad de la pelota, y calcular luego la fuerza neta que actúa sobre ella, lo cual nos dará su aceleración. Imaginamos que la pelota permanece siempre sobre la superficie del paisaje, y no volando sobre él, ni enterrada o semienterrada por debajo de la superficie. Para mayor simplificación, consideraremos una sola dimensión x para el movimiento. La pelota se mueve asimismo en una segunda dimensión, la altura, pero convendremos que eso será la altura del paisaje, h(x). (No permitiremos a nuestra pelota ganar tiempo con saltos en suspensión).

La fuerza neta es la suma de fuerzas diferentes: la fuerza de gravedad que tira hacia abajo y la fuerza ejercida por el propio terreno. La fuerza de la superficie del terreno se llama fuerza normal, no porque sea especialmente corriente, sino porque es siempre perpendicular a la inclinación de la pendiente, por lo que aquí «normal» es simplemente sinónimo de «perpendicular» (como lo sería «ortogonal», en este contexto).

Considere el caso más simple que podamos imaginar, el de un paisaje llano, de tal manera que la altura h(x) es la misma sea cual sea el valor de x. La fuerza de gravedad g tira hacia abajo, mientras que la fuerza normal n es perpendicular al suelo, por lo que en este caso empuja directamente hacia arriba. Sabemos por tanto al instante que no hay fuerza neta a izquierda o derecha, ya que la fuerza neta es la suma de una fuerza que tira directamente hacia abajo y otra directamente hacia arriba. Y como hemos

dicho que la pelota no va a hundirse en el suelo ni a saltar por el aire, la fuerza neta será exactamente igual a cero. En otras palabras, la fuerza de gravedad y la fuerza normal son iguales y opuestas, y al sumarse no añaden ninguna cantidad de fuerza; por consiguiente, la pelota no experimenta aceleración. Esto es así sea cual sea la velocidad inicial de la pelota; y sea cual sea aquello que la pone en movimiento, seguirá moviéndose en la misma dirección a velocidad constante.



Cuando el perfil del paisaje es en pendiente, las cosas se vuelven más interesantes. La gravedad sigue empujando a la pelota en línea recta hacia abajo, pero ahora la fuerza normal no será vertical. Como resultado, cuando sumamos ambas fuerzas, el valor de la fuerza neta resultante no será igual a cero. La fuerza total, y, por tanto, la aceleración, tendrá una dirección «cuesta abajo». Si partimos de una situación en que la pelota está en reposo, comenzará a rodar hacia abajo; si en la situación de partida ya se estaba moviendo, cobrará velocidad; y si rueda cuesta arriba, se ralentizará.

Nuestra pelota rodante tiene tanto energía potencial como energía cinética. La energía potencial es:

$$V(x) = mgh(x) \qquad (3.2)$$

donde m es la masa de la pelota, g = 9,8 m/s² es la aceleración debida a la gravedad, y h(x) es la altura de la colina en cada punto x. La expresión puede ser un poco confusa si no está acostumbrado a los signos de multiplicación implícitos; m es un número que corresponde a una constante, g también lo es, y h(x) es una función que asigna un número h a cada valor de x. Multiplicamos los tres números para obtener la energía potencial, que convencionalmente se expresa como V, aunque en ocasiones se utilizan otras letras. Dado que es la energía por la que vamos a interesarnos, y no por la altura de la colina, solemos hablar directamente en términos del potencial V(x).

Si quisiéramos servirnos de la segunda ley de Newton, , para calcular el movimiento de la pelota, esta función de la energía potencial nos proporciona una buena manera de abordar la cuestión. La pelota real se mueve tanto vertical como horizontalmente, a tenor de las ondulaciones del terreno, así que la idea es sumar vectores para obtener la fuerza neta, tal como hemos hablado. Pero puesto que estamos imaginando que la pelota permanece en contacto con el suelo durante todo el tiempo, en realidad es únicamente el movimiento horizontal lo que necesitamos calcular; las subidas y bajadas en vertical son secundarias.

Así que tenemos que pensar en la aceleración en dirección x, que procede de la fuerza en dirección x. Intuitivamente, esperamos que esta fuerza provenga de la pendiente del potencial: la pelota será impelida cuesta abajo, y cuanto más pronunciada sea la pendiente, mayor el impulso. En términos de cálculo, la fuerza en dirección x es menos la derivada del potencial con respecto a la posición:

$$F_x = -\frac{dV}{dx} \qquad (3.3)$$

La expresión Fx nos recuerda que se trata de la fuerza en dirección x. Y el signo es de mero sentido común: un potencial cuesta arriba (dV/dx positivo) empuja a la pelota hacia la izquierda, en correspondencia con una fuerza negativa en dirección x.

DESDE LA PERSPECTIVA DE LA ENERGÍA

Una vez que conocemos la fuerza ejercida sobre la pelota, podemos arremangarnos y resolver con exactitud cómo se desenvolverá la partícula en cualquier paisaje especificado, utilizando la segunda ley de Newton para calcular la aceleración y haciendo luego algunas integrales. Solo que todo esto suena un poco tedioso. Así que recurramos mejor a la intuición para vislumbrar un poco por dónde irán los tiros, a partir de la noción de conservación de la energía.

Además de la energía potencial, teníamos la energía cinética:

$$E_{\text{cinética}} = \frac{1}{2}mv^2 \qquad (3.4)$$

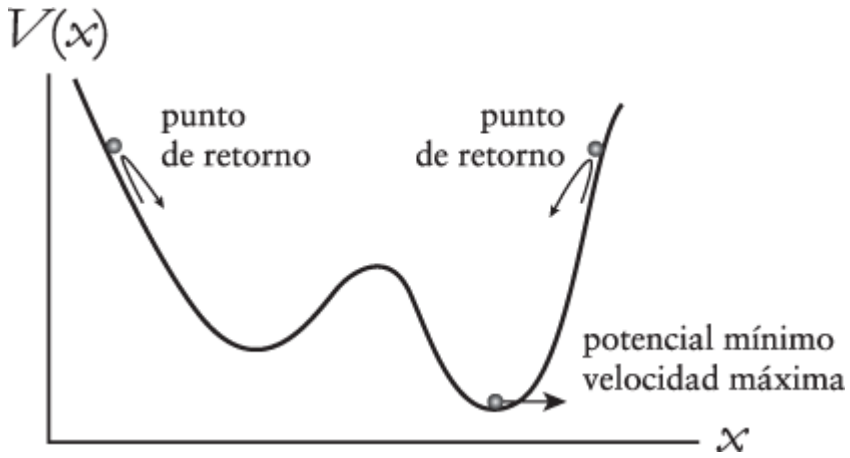
donde v es la velocidad de la pelota. El número obtenido al sumarlas, la energía total, permanece constante a lo largo de la trayectoria de la pelota:

$$E_{\text{total}} = V + E_{\text{cinética}} \qquad (3.5)$$

Esta simple ecuación nos dice por sí misma bastante acerca del comportamiento de la pelota. La energía cinética se reduce al mínimo posible (es decir, a cero) cuando la velocidad desaparece. Sabemos así de inmediato que la energía potencial, y, por tanto, la altura de la pelota, alcanza su nivel máximo cuando la velocidad es cero. De forma similar, la velocidad será máxima cuando la energía potencial esté al mínimo, al pie de la pendiente.

La conservación de la energía nos ofrece una buena visión de conjunto de la trayectoria de la pelota. Imagine que soltamos la pelota, con velocidad

inicial cero, en el fondo de un valle. Como no se ejerce ninguna fuerza sobre la pelota, se queda quieta donde está. Lo cual concuerda perfectamente con nuestra intuición, según la cual si dejamos una pelota en reposo en el fondo de un valle no se moverá.



Imagine ahora que la pelota parte de nuevo desde un estado de reposo, pero esta vez en porción en pendiente de la colina. La energía total será simplemente la energía potencial en esa ubicación. La pelota rodará pendiente abajo, cobrará velocidad e irá convirtiendo energía potencial en energía cinética. Si el perfil del paisaje comienza a ser cuesta arriba en algún punto, la pelota perderá velocidad a medida que gana energía potencial. Y si la pelota llega a alcanzar la misma altura desde la que comenzó a rodar, su velocidad volverá a ser cero, puesto que habrá recuperado exactamente la misma energía potencial con la que partió. A esto se lo conoce como punto de retorno en la trayectoria, ya que la pelota volverá a rodar hacia abajo a partir de ese punto en la dirección de la que vino.

En este mundo ideal de fricción cero, la pelota se desplazará por siempre hacia delante y hacia atrás entre los dos puntos de retorno. Podríamos pensar que al llegar al fondo en sus idas y venidas llegará a detenerse en alguna ocasión, pero este es uno de los casos en que la intuición nos engaña: estamos acostumbrados a un mundo con fricción. Si la energía de la pelota se conserva estrictamente, oscilará en una dirección y en otra por toda la eternidad.

EL OSCILADOR ARMÓNICO SIMPLE

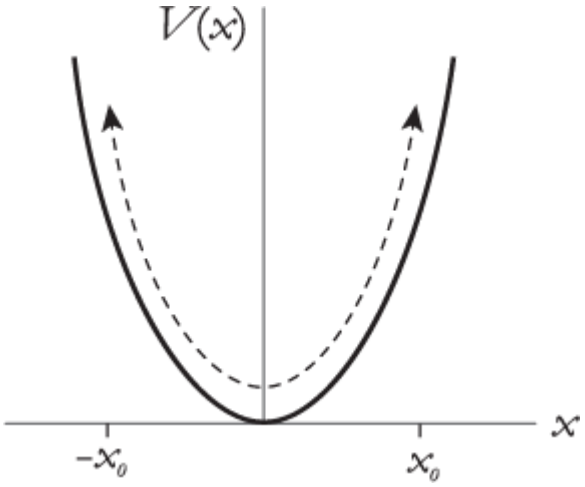
Hay vacas esféricas y vacas esféricas. La vaca esférica predilecta de toda la física —el sistema físico simple y de solución precisa más importante, de aplicación asombrosamente amplia— es el oscilador armónico simple.

Consideremos de nuevo nuestra pelota en la colina carente de fricción, pero seamos ahora un poco más concretos a la hora de configurar el perfil del paisaje. Este tendrá forma de valle parabólico, con el valor mínimo en x = 0, de tal manera que la energía potencial será:

$$V(x) = V_0 x^2 \qquad (3.6)$$

Aquí V0 es un parámetro que nos dice lo amplia (si V0 es pequeño) o delgada (si V0 es grande) que es la parábola.

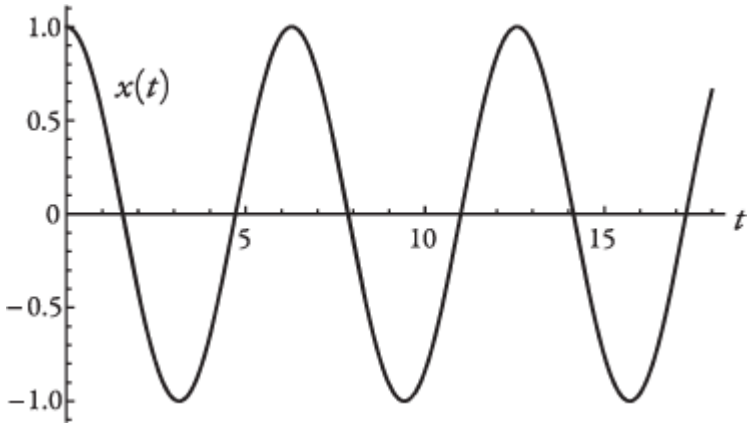
Observando simplemente el potencial, podemos fácilmente determinar lo que va a hacer la pelota. Si la soltamos con velocidad cero en un punto cualquiera del brazo derecho, digamos x = x0, comenzará a rodar hacia el fondo y luego subirá por el brazo izquierdo. Dado que el potencial es simétrico en torno a x = 0, por la conservación de la energía sabemos de inmediato que la pelota remontará hasta x = −x0, donde la energía potencial volverá a tener el mismo valor que en el punto de partida. Ahí, la pelota tendrá de nuevo velocidad cero y volverá a rodar siguiendo el camino de retorno. Regresará hasta x = x0, y en este punto el ciclo volverá a comenzar, y seguirá repitiéndose por siempre.



Podríamos pensar en términos de fuerzas, en lugar de hacerlo como hemos hecho en términos de energías. De (3.3) se desprende que la fuerza en dirección x es menos la derivada del potencial. Si el potencial es $V(x) = V0x^2$, y por (2.9) sabemos que $d(x^2)/dx = 2x$, tenemos entonces que $Fx = -dV/dx = -2V0x$. De modo que cuando x es negativa, la fuerza es positiva, e impele a la partícula de vuelta hacia la derecha; cuando x es positiva, la fuerza la impulsa hacia la izquierda. En cualquier caso, la fuerza impulsa a la partícula de vuelta hacia el mínimo en $x = 0$. Se llama a esto fuerza restauradora, y su magnitud es proporcional al desplazamiento $x = 0$ del punto de equilibrio.

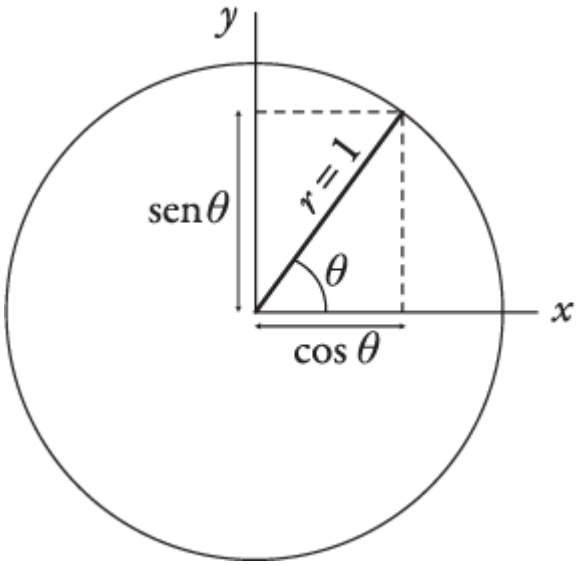
Este tipo de sistema se llama «oscilador», debido a su movimiento de vaivén. Es «armónico» cuando el potencial es exactamente proporcional a x^2 (un potencial x^4 sería un oscilador, pero no un oscilador armónico), y es «simple» cuando la energía se conserva íntegramente (por no haber fricción). Hay también osciladores armónicos «amortiguados», con fricción no cero, y osciladores «forzados», cuando aplicamos cada vez más energía al sistema.

Podemos representar gráficamente el comportamiento de un oscilador armónico simple. Establecemos $x0 = 1$ para simplificar; la partícula parte de $x = 1$, desciende hasta -1, oscila de vuelta hasta 1, y continúa siguiendo este patrón.

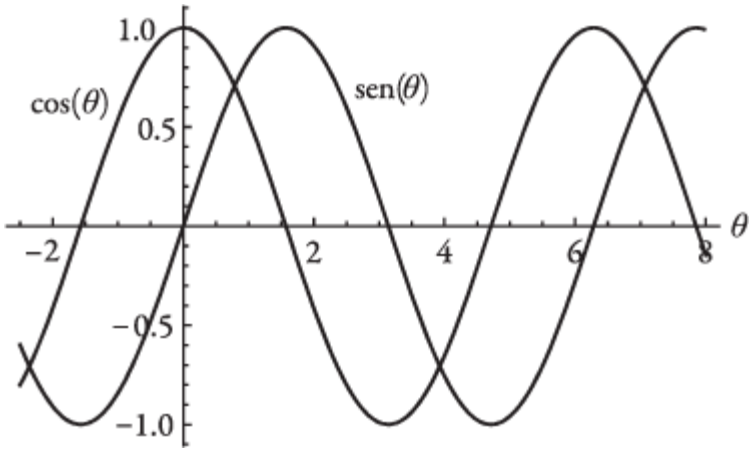


Este tipo de funciones pueden resultarle familiares (o activarle un flashback) si alguna vez estudió un poco de trigonometría. Hay dos funciones trigonométricas especialmente importantes: el seno, que parte de cero, sube hasta +1, baja a −1 y regresa; y el coseno, que parte de +1, pasa por 0 hasta llegar a −1 y vuelve a subir de nuevo.

La forma más fácil de definir estas funciones trigonométricas es pensar en una circunferencia unitaria, es decir, una circunferencia de radio 1. Cualquier punto que tomemos de la circunferencia está especificado únicamente por un ángulo θ del eje x. Mediremos este ángulo en radianes, unidades por las cuales 360 grados se corresponden con 2π radianes, donde π = 3,14159… es la famosa constante numérica igual a la circunferencia de un círculo dividida por su diámetro. Así pues, un radián = 180/π grados. Hay un buen número de razones para esta elección, la más importante de las cuales es que las derivadas e integrales de los senos y cosenos asumen elegantes formas cuando medimos ángulos usando radianes. En estas unidades, cosθ es la proyección en vertical del punto elegido de la circunferencia hasta el eje x, y senθ es la proyección lateral al eje y.



Podemos observar que cos (0) = 1 y sen (0) = 0. A partir de ahí, ambas funciones oscilan hacia arriba y hacia abajo según el valor del ánguloθ va de 0 a 2π radianes.



Si comparamos la representación de estas funciones trigonométricas con la evolución del oscilador armónico simple, vemos que la posición del oscilador se parece mucho a la función coseno. Cosa que resulta cierta con toda precisión. Para un oscilador general que parte del estado de reposo, la posición evoluciona como:

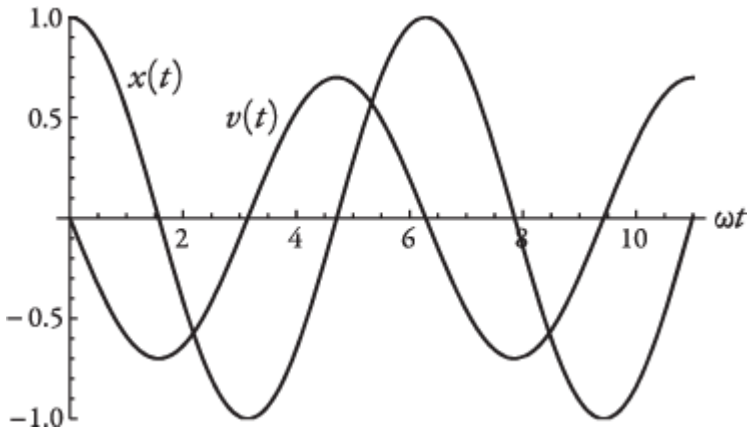
$$x(t) = x_0 \cos(\omega t) \qquad (3.7)$$

El símbolo ω corresponde a la letra griega omega, que expresa la frecuencia angular del oscilador. Siempre que se nos presente un fenómeno oscilatorio, la frecuencia f es la razón a la que el oscilador retorna a su punto de partida, mientras que la frecuencia angular ω es la razón con que el ángulo pasa de 0 a 2π; la relación entre ambas se expresa: ω = 2πf. Para el oscilador armónico potencial (3.6), la frecuencia angular es: ω = √2V0.

Podemos plantearnos ahora la velocidad del oscilador. Parte de cero, puesto que soltamos nuestra partícula estando en reposo. La partícula comienza a moverse hacia la izquierda, de modo que la velocidad se convierte en negativa. Al llegar al punto de retorno, la velocidad vuelve a cero, y a partir de ahí va oscilando de un lado a otro. Todo esto suena un poco a función seno, solo que al revés (ya que senθ parte de cero y sube, mientras que v(t) parte de cero y baja). También esto es correcto con toda precisión:

$$v(t) = -v_0 \operatorname{sen}(\omega t) \qquad (3.8)$$

La velocidad oscila con la misma frecuencia angular que la de la posición. El coeficiente v0 dependerá de la masa de la partícula representada por el oscilador; puede relacionarla con x0 y V0 si piensa en la conservación de energía.



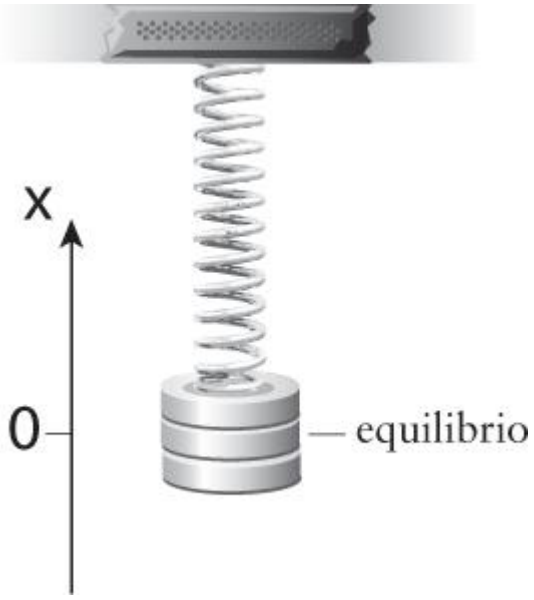
OSCILADORES ARMÓNICOS POR DOQUIER

Poder resolver ecuaciones de movimiento con exactitud, como es el caso del oscilador armónico simple, es tan satisfactorio como útil. Nos proporciona un sentimiento de haber alcanzado todo un logro. Y es una capacidad bastante rara en situaciones físicas reales. Ni siquiera un modesto potencial de cuarta potencia como $V(x) = V0x^4$ tiene soluciones exactas expresables en términos de simples funciones. Aunque solo fuera por esta razón, el oscilador armónico sería ya muy valioso.

Y lo mejor es comprobar cómo tu sistema de soluciones exactas se presenta una y otra vez en el mundo real. Por fortuna, eso es justamente lo que sucede con el oscilador armónico simple.

Piense en otro tipo diferente de sistema físico, sin relación aparente con ninguna pelota rodando por ninguna colina. Vamos a imaginar un peso colgado de un muelle y suspendido desde lo alto. Si tiramos del peso de tal forma que el muelle se estire, el muelle ejercerá una fuerza opuesta hacia arriba. Pero igualmente es verdad —al menos en un mundo ideal en que los muelles son perfectos y no se doblan ni se enredan— que si empujamos el peso hacia arriba, el muelle ejercerá una fuerza opuesta hacia abajo. Habrá un punto de equilibrio en medio, un punto en el que al soltar el peso las fuerzas estén equilibradas y el peso no se mueva. Si lo desplazamos un

poco hacia arriba o hacia abajo con respecto a ese punto de equilibrio, el peso oscilará arriba y abajo.



Ha resultado que el movimiento vertical de nuestro sistema peso-muelle describe una onda sinusoidal, ni más ni menos que como nuestra pelota en un potencial parabólico. (Los senos y los cosenos, junto con cualquier otra versión modificada a partir de los mismos, suelen llamarse «ondas sinusoidales» o simplemente «sinusoides»). Vale la pena demorarnos un poco en esto. Los sistemas físicos en estos dos casos son completamente independientes uno de otro: tenemos una pelota rodando de derecha a izquierda por un valle, por un lado; y un peso colgado de un muelle por otro. Pero las ecuaciones implícitas son equivalentes. Desde la perspectiva abstracta de un físico teórico, son el mismo sistema. (Un físico experimental que tuviera que construir el sistema podría no estar de acuerdo).

Hay una razón profunda por la que el oscilador armónico simple es tan ubicuo en física: hay un enorme número de sistemas, incluidos los campos cuánticos vibratorios del modelo estándar en física de partículas, que son aproximadamente osciladores armónicos simples. No es difícil entender por qué.

Consideremos un sistema físico que oscile de un lado a otro sin fricción, de forma que la energía se conserva. Habrá un punto de equilibrio con respecto al cual oscila el sistema (o en el cual permanece quieto, si el sistema parte de un estado de reposo en ese punto). Sea x la cantidad que mide lo lejos que se desplaza desde ese punto de equilibrio. La energía potencial es una función V(x). Por el momento imaginaremos que esta función es completamente arbitraria.

He aquí ahora un hecho matemático crucial: podemos expresar nuestra función de energía potencial como una serie infinita, como una suma de términos cada uno de los cuales adopta la forma x elevada a una potencia:

$$V(x) = a + bx + cx^2 + dx^3 + ex^4 + \ldots \qquad (3.9)$$

Eligiendo cuidadosamente los coeficientes numéricos {a, b, c…}, cualquier función que se comporte puede representarse bajo esta forma. («Que se comporte» excluye cosas como una función cuyos valores dieran saltos de discontinuidad).

Pensemos un poco en esta expresión. El primer parámetro, a, es tan solo un número constante. No tiene ningún efecto en la pendiente del potencial; con cada valor diferente de a se corresponde un desplazamiento del potencial hacia arriba o hacia abajo, sin modificación de su forma. Pero es solo la pendiente del potencial lo que crea una fuerza, no el valor numérico del potencial. Por (3.3) sabemos que la fuerza sobre el sistema es la derivada del potencial, que no tiene en cuenta a. Así, podemos establecer a = 0, sin que cambie el comportamiento del sistema de modo relevante en ningún sentido.

Pensemos ahora en el resto de la serie infinita. Hemos definido x de tal forma que x = 0 en el punto de equilibrio, donde el sistema puede quedarse tal cual está, sin moverse. ¿Qué sucede si agitamos el sistema levísimamente? Cada toquecito mínimo se correspondería con valores de x muy pequeños. Y cuando tomamos un número pequeño (mucho menor que uno) y lo multiplicamos por sí mismo, obtenemos un número aún más

pequeño. Así que, considerando todos los términos de (3.9), asumen un valor decreciente a medida que avanzamos en la serie asignando un valor creciente a las potencias de x. Para todo valor lo suficientemente pequeño de x, lo único que debería contar es el primer término, bx. Esto es una aproximación, por supuesto, pero una aproximación que es cada vez mejor a medida que consideramos valores cada vez más pequeños de x. No importa cuáles sean los demás coeficientes, siempre habrá valores de x tan pequeños, que solo importará el primer término de la serie.

Pero, un momento. Si x = 0 es el punto de equilibrio, eso significa que V(0) tiene que ser un valor mínimo —el fondo del valle—, donde la pendiente es cero y no hay fuerza alguna. Pero en la aproximación de x pequeña, en que V(x) bx, la pendiente de V en x = 0 es simplemente b. No es cero a menos que la propia b sea cero. Así pues, dados todos estos supuestos, podemos establecer b = 0 de la misma forma que establecíamos a = 0; dicho de otro modo, x = 0 no es un mínimo en absoluto. El potencial que nos queda es:

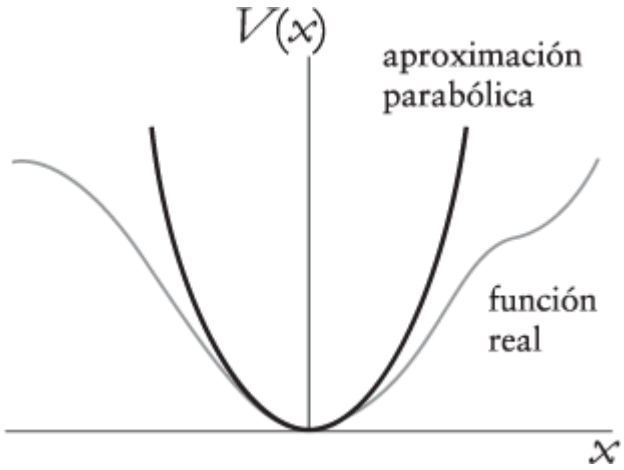
$$V(x) = cx^2 + dx^3 + ex^4 + \dots \qquad (3.10)$$

Y ahora podemos repetir el mismo argumento de antes: para valores suficientemente pequeños de x, los términos de orden superior se vuelven irrelevantes. En otras palabras, en términos extremadamente generales sabemos que la forma del potencial para un sistema oscilante en las proximidades del estado de equilibrio asume la forma aproximada:

$$V(x) \approx cx^2 \qquad (3.11)$$

Esto no es más que una parábola, el potencial para un oscilador armónico simple. Es un resultado maravilloso: para pequeñas desviaciones del estado de equilibrio, en ausencia de fricción, a casi todo sistema oscilatorio se aproxima un oscilador armónico simple.[2] Nuestro «oscilador» puede ser una pelota rodando cerca del fondo de un potencial, o un peso colgado de una muelle cerca de su punto de equilibrio; pero podría ser también algo como

el desplazamiento de un péndulo, o un átomo agitándose en una molécula, o la amplitud de una onda de sonido, o la corriente eléctrica de un circuito, o el valor del campo de bosones de Higgs. Mientras sea un sistema que pueda describirse en términos de energía cinética y potencial, puede expresarse en forma de parábola la aproximación del potencial a su mínimo, por cuanto el comportamiento físico del potencial es aproximadamente el de un oscilador armónico. Generalmente la aproximación no será exacta; una vez comencemos a tener en cuenta mayores desviaciones a partir del estado de equilibrio, todos esos otros términos de (3.10) empezarán a cobrar relevancia. Pero estas son complicaciones que podemos añadir más tarde.



Esta discusión aporta un poco de sustancia a la filosofía de la vaca esférica. No es tan solo cuestión de considerar sistemas simplificados al máximo y esperar tener suerte. La idea de tomar una expresión complicada y desarrollarla en forma de suma de términos infinitos, como en (3.9), es una técnica aplicable a una amplia variedad de problemas. A menudo somos lo bastante afortunados como para que los términos según avanzan en la serie sean numéricamente más pequeños que los primeros. Esto nos permite desarrollar un procedimiento sistemático conocido como teoría de perturbaciones: exponemos las ecuaciones que rigen el sistema como la suma de algo simple más una pequeñísima perturbación, resolvemos con exactitud la parte simple y a partir de ahí vamos añadiendo el resto paso a paso. A veces (no siempre) el universo nos ayuda a seguir adelante en nuestro empeño por entenderlo.

ESPACIO DE FASES

De acuerdo con el paradigma laplaciano, la trayectoria de un sistema se determina a través de las posiciones y velocidades de cada una de las partes del sistema en cierto instante. Y hemos dicho que el momento de un objeto es el producto de su masa por su velocidad, . Dado que habitualmente consideramos sistemas en los cuales la masa es constante, en términos de la información que necesitamos para predecir el comportamiento de un sistema, especificar «posición y momento» es equivalente a especificar «posición y velocidad». Una vez hayamos entrado en temas más avanzados de la física —tan pronto como el capítulo siguiente—, veremos por qué el momento es más fundamental que la velocidad desde cierta perspectiva, por lo que generalmente es lo que utilizaremos.

Recapitulando, el conjunto de todas las posibles posiciones y momentos en un sistema se conoce como espacio de fases de ese sistema.

$$\text{espacio de fases} = \{\vec{x}_i, \vec{p}_i\} \text{ para todos los objetos } i$$

(Las llaves {} suelen utilizarse para definir conjuntos). Especificar en qué punto se halla el sistema en el espacio de fases en cualquier momento es tanto como determinar su evolución entera de acuerdo con la mecánica newtoniana; en otras palabras, el espacio de fases es el conjunto de todos los posibles estados en que puede estar el sistema.

A modo de simple indicio de por qué el momento es más básico que la velocidad, piense en la segunda ley de Newton, . Sabemos que la aceleración es la derivada de la velocidad con respecto al tiempo. Y esta formulación del principio de Newton asume implícitamente que la masa no cambia con el tiempo, es simplemente una constante. Por tanto, podemos pensar en m no solo como «masa por derivada de la velocidad», sino

también, de forma equivalente, como «la derivada de masa por velocidad», ya que una constante puede ponerse dentro o fuera de la derivada:

$$m\frac{d}{dt}\vec{v} = \frac{d}{dt}\left(m\vec{v}\right) \text{ [cuando } m \text{ es constante]} \qquad (3.12)$$

Esto sugiere una forma elegante de rescribir la segunda ley de Newton en términos de la derivada del momento:

$$L(x, v) = \frac{1}{2}mv^2 - V(x) \qquad (3.13)$$

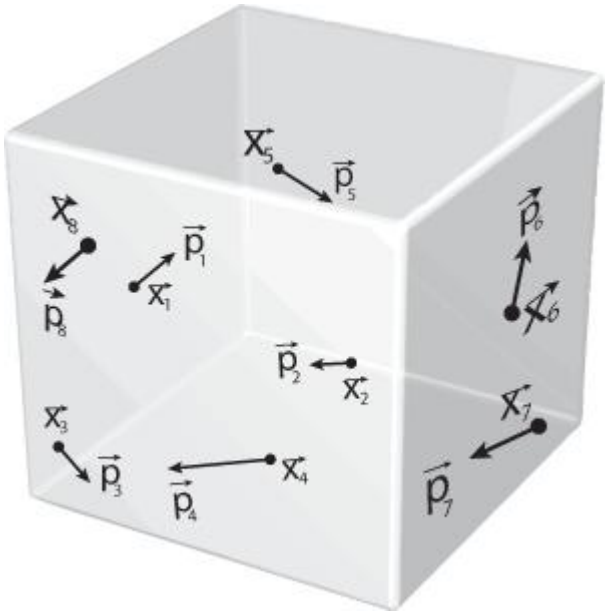
Esta expresión no solo es más compacta, sino también más general que , por cuanto esta forma sigue siendo válida incluso cuando la masa del objeto cambia (por ejemplo, a medida que un cohete pierde masa paulatinamente según expele los gases de combustión). La fuerza es la razón a la cual el momento cambia con el tiempo.

En el mundo de nuestra experiencia, los objetos ocupan un lugar en un espacio tridimensional. Los matemáticos, y los físicos a su estela, se han replanteado la palabra «espacio» para darle un significado mucho más general: básicamente, sería cualquier disposición con alguna estructura adicional. Entonces, «el conjunto de todas las posibles posiciones de un simple objeto» vale para el viejo espacio tridimensional. Pero el espacio de fases para ese objeto sería hexadimensional, con tres dimensiones para la posición y otras tres para el momento (que es un vector tridimensional).

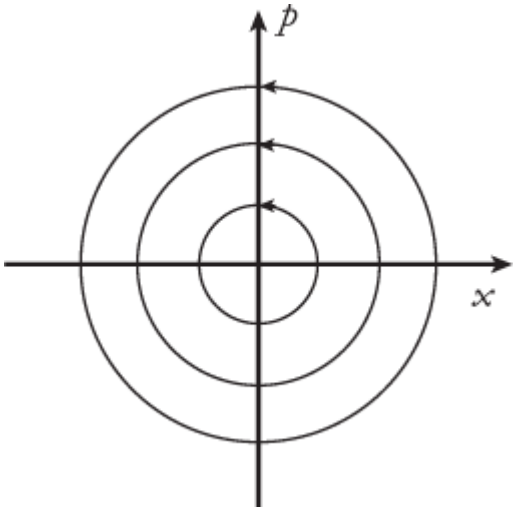
Si nuestro sistema consiste en N partículas en el espacio tridimensional, podemos hablar del espacio de configuración del sistema, un espacio de 3N dimensiones definido por la posición tridimensional de cada partícula. Habrá un correspondiente espacio de fases de 6N dimensiones, ya que cada partícula tendrá también un momento tridimensional. El sistema Tierra-luna —dos objetos moviéndose en un espacio tridimensional— tiene un espacio

de fases dodecadimensional. Por no mencionar que si nuestros objetos son más complicados que simples partículas, podríamos necesitar especificar también sus orientaciones y momentos angulares.

En general, el espacio de fases puede ser una construcción matemática altamente multidimensional, dependiendo del tipo de sistema en que estemos pensando. La idea fue explorada durante el siglo XIX por Ludwig Boltzmann, James Clerk Maxwell y otros pioneros de la mecánica estadística, quienes regularmente contemplaban sistemas compuestos por muchas partículas. Una cuantificación tradicional de «muchas» es el número de Avogadro, 6×10^{23}, que representa aproximadamente el número de átomos contenido en un gramo de hidrógeno monoatómico. Un gramo es pequeño pero perceptible a escala humana, de modo que cualquier cosa de ese tamaño o mayor es tranquilizadoramente macroscópico. Una colección con un número de Avogadro de partículas se describe como un espacio de fases de $3{,}6\times10^{24}$ dimensiones. Es bastante grande.

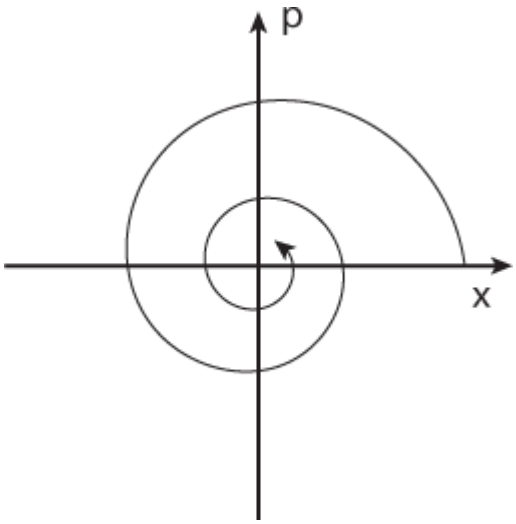


Pero un espacio de fases puede también ser bastante modesto. El oscilador armónico simple tiene un espacio de configuración unidimensional, definido por su posición x, de modo que su espacio de fases es bidimensional, {x, p}. Esto hace posible visualizarlo, y es útil hacerlo.



En esta figura hemos representado tres posibles trayectorias de fases para un oscilador armónico simple particular. La posición oscila de un lado a otro, y lo mismo sucede con el momento, pero con diferentes puntos de partida: cuando la posición está en cero, el momento alcanza su máximo, y viceversa. Así, cualquier posible trayectoria del sistema describe una órbita elíptica en un espacio de fases, con elipses de diferentes tamaños que se corresponden con condiciones iniciales diferentes. (Es una elipse en un espacio de fases, no en el espacio real; en el espacio real, el oscilador simplemente se mueve de un lado a otro. El espacio de fases expresa una idea diferente. No necesitamos especificar independientemente la «velocidad en el espacio de fases»; la trayectoria entera queda determinada al saber de qué punto partimos). La frecuencia del oscilador es independiente de su amplitud inicial, así que el sistema emplea la misma cantidad de tiempo en recorrer el perímetro de un círculo, sin importar el tamaño de este.

Por mera diversión, preguntémonos qué sucedería si añadiéramos un poco de fricción al sistema, para tener un oscilador armónico amortiguado en lugar de uno simple. Intuitivamente sabemos lo que pasará: el oscilador se balanceará de un lado a otro, pero irá perdiendo energía, por lo que el balanceo disminuirá en amplitud de forma continuada. En el diagrama de fases, esto se representaría sustituyendo nuestras elipses por espirales desplazándose hacia el centro.



EL ESPACIO DE CAMINOS

No está mal lo que hemos hablado de pelotas rodando por colinas y osciladores balanceándose de aquí para allá. Vamos a recompensarnos con algo de verdad alucinante.

Al explorar la mecánica clásica, hemos hecho repetido hincapié en el paradigma laplaciano, según el cual especificamos el estado del sistema —un punto en un espacio de fases, con la posición y el momento de cada parte del sistema en un instante del tiempo— y eso determina su trayectoria entera. Volviendo al comienzo de este capítulo, eso casi es tanto como ponernos una venda en los ojos y caminar en línea recta. Sabemos lo que estamos haciendo en este momento, y a partir de ahí podemos avanzar paso a paso en el tiempo, de un instante al siguiente. Los físicos denominamos a este tipo de procedimiento un problema de valor inicial, puesto que partimos de ciertas condiciones iniciales y elaboramos a partir de ellas todo lo demás.

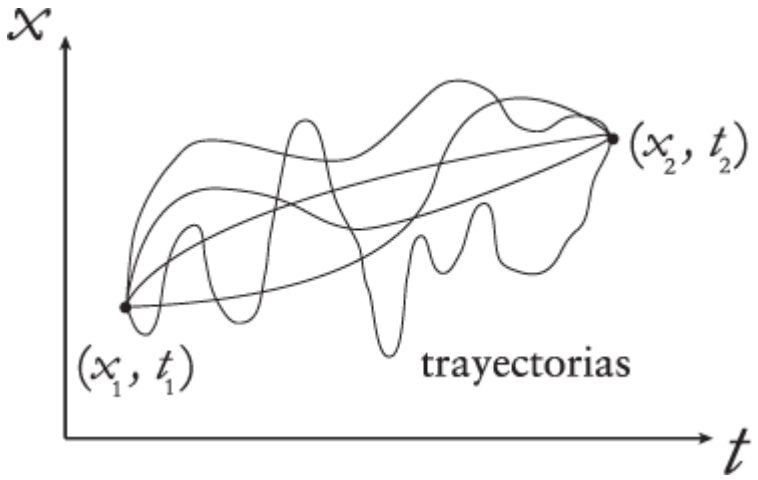
Pero hay otras formas de trazar una línea recta, como tensar una cuerda entre dos árboles. En este caso, no necesitamos pensar en ningún tipo de «dirección inicial». Tan solo necesitamos los dos árboles, y en cuanto

tendemos la cuerda tensándola bien entre ambos, tenemos automáticamente una línea recta que apunta en la dirección correcta. Ya no se trata de un problema de valor inicial, sino de un problema de valor de frontera, en el que tenemos una información al comienzo (la localización del primer árbol) y otra información al final (la ubicación del otro árbol).

Podemos describir la cuerda tendida entre dos árboles en términos de propiedades globales, en lugar de locales. (Recuerde la distinción entre la teoría de las órbitas planetarias de Kepler y la de Newton). «Caminar en la misma dirección» es un procedimiento pura y esencialmente local: hacemos algo específico en cada momento. Mientras que «obtener la distancia más corta» es una propiedad global: necesitamos comparar la longitud de la cuerda entera con la habría tenido de no ser recta.

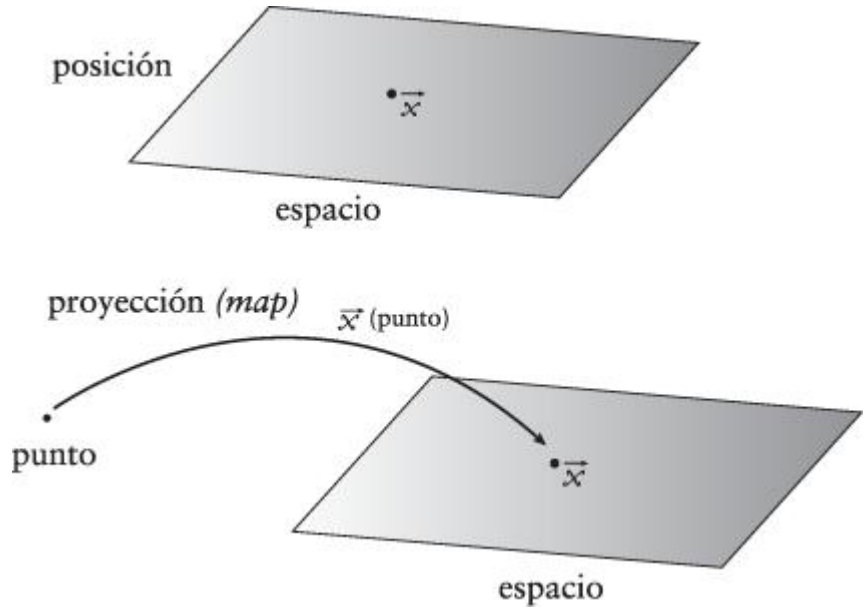
Lo que estamos haciendo implícitamente es imaginar un espacio matemático inmensamente grande, el espacio de todas las formas posibles en que una cuerda podría unir un punto con otro. La mayoría de tales posibles formas no serían en modo alguno rectas: hay un único camino recto, y un número infinito de caminos curvilíneos. De entre todo ese gigantesco conjunto de posibilidades, el camino recto se distingue por tener una longitud más corta que cualquiera de todos los demás.

Lo más destacable es que toda la mecánica clásica puede migrarse a este tipo de lenguaje global, como alternativa a la perspectiva local de progresivos avances en el tiempo que hemos estado adoptando hasta ahora. En lugar de especificar la posición y el momento de una partícula en un instante del tiempo, imagine que especificamos simplemente la posición x1 en cierto instante inicial t1, y también la posición x2 en un instante posterior t2. Así como toda curva en el espacio tiene una longitud, toda trayectoria concebible de una partícula entre (x1, t1) y (x2, t2) tiene (como veremos) una cantidad llamada acción, que depende del modo en que la energía cinética y la energía potencial evolucionan a lo largo del camino. De entre todas las trayectorias posibles que la partícula podría seguir, la que realmente sigue —la trayectoria que obedece a las leyes de Newton— resulta ser la trayectoria con la acción más pequeña.

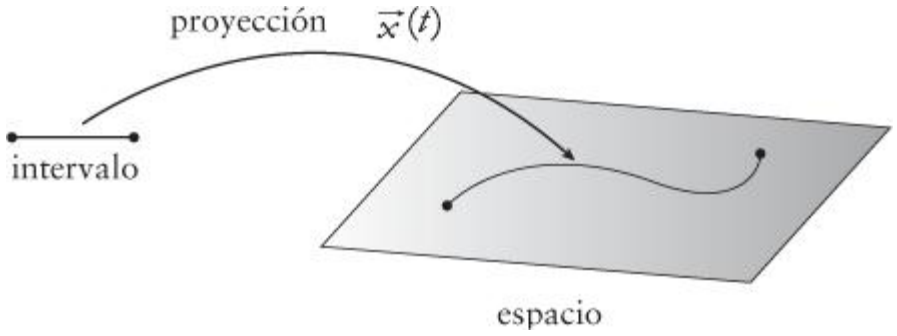


Esta idea se conoce, con buen criterio, como principio de mínima acción, el cual se desarrolló gradualmente durante el periodo de tiempo transcurrido entre los siglos XVII y XIX. El principio de mínima acción presupone partir desde estas nociones sobre las que hemos estado pensando, de modo que vamos a detenernos a reflexionar acerca de la filosofía matemática que hay detrás.

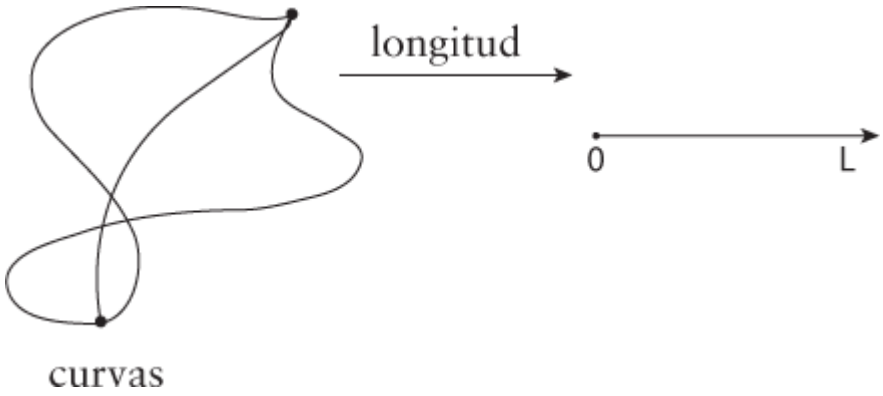
Las matemáticas, en gran parte, pueden abordarse estudiando los espacios y las proyecciones entre ellos. Una función simple f(x), después de todo, es una proyección entre el conjunto de números reales consigo mismo. Ya hemos estado recurriendo implícitamente a muchos otros ejemplos. Cuando decimos «la posición de una partícula», lo más normal es que visualicemos un punto literalmente incrustado en el espacio. Pero un matemático conceptualizaría esto mismo como una proyección de un simple punto al espacio tridimensional.



De forma similar, la trayectoria de una partícula en el espacio a través del tiempo puede pensarse como una proyección de un intervalo al espacio.



Y por lo mismo, la noción de longitud de una curva puede expresarse estableciendo una proyección entre el espacio de todas las curvas (digamos, entre dos puntos determinados) y los números reales no negativos, por la cual a cada curva se le asocia su longitud L.



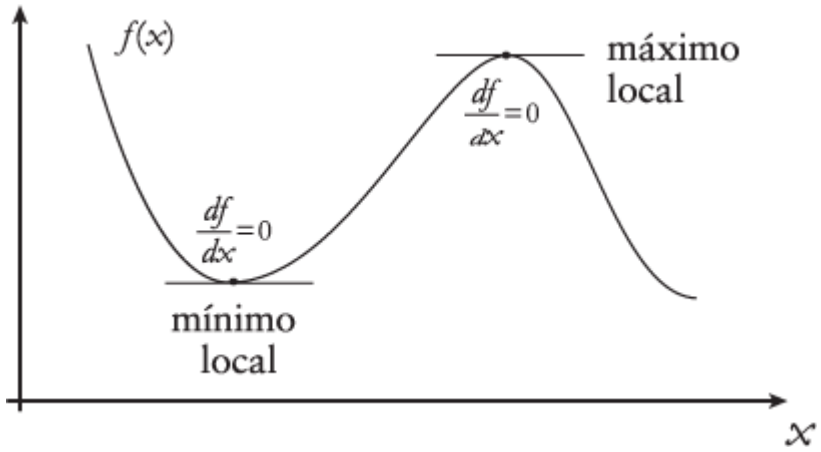
Cabe admitir que «el espacio de todas las curvas entre dos puntos determinados» puede ser difícil de visualizar. Por eso intentamos esbozar unos pocos ejemplos representativos y esperamos que a partir de ellos usted pueda imaginar el resto.

Un punto es un espacio de dimensión cero, una línea es unidimensional, un plano es bidimensional, y así sucesivamente. Hay una dimensión para cada nueva información que usted necesite para especificar en qué lugar del espacio se encuentra. El espacio de todas las curvas es, por tanto, de dimensión infinita, ya que especificar una curva es especificar un infinito número de puntos en su recorrido. Esto crea algunas sutilezas matemáticas interesantes, pero lo más notable es lo competente que es nuestra caja de herramientas matemáticas corriente en espacios de dimensión infinita sin necesidad de muchas modificaciones.

En particular, existe un modo de cálculo para espacios de dimensión infinita llamado cálculo de variaciones. No vamos a ahondar en él, por divertido que pudiera ser. La vida es breve, y tenemos por delante un buen montón de grandes ideas. Pero quiero hacer mención de él porque el cálculo desempeña un papel crucial en la idea de «encontrar la curva de longitud mínima».

Volvamos a considerar una función simple de una variable, f(x). La derivada df/dx nos proporciona la pendiente de esa función en cada punto. Solo mirando la gráfica, advertimos una propiedad importante: en cada uno de los puntos en que la función alcanza un máximo local (lo más alto de una

colina) o un mínimo local (el fondo de un valle), la derivada es exactamente cero.



Aquí radica el secreto matemático que se oculta tras el hecho de encontrar curvas de longitud mínima, o cualquier otra propiedad. Consideremos el conjunto de todas las curvas posibles. Podemos definir una función sobre este conjunto (una correspondencia entre cada curva y un número), dada por «la longitud de la curva». Tomemos luego la derivada de esa función con respecto a todas las formas posibles en que podemos deformar infinitesimalmente la curva. Si llegamos al mínimo, todas esas derivadas desaparecerán con total precisión. Así pues, podemos transformar la idea verbal «camino más corto» en un conjunto de ecuaciones matemáticas: «desaparición de la derivada de la función longitud sobre el espacio de caminos».

MÍNIMA ACCIÓN

Pero no es la longitud de una curva lo que nos interesa en este momento. Cuando tiramos una piedra y esta sigue cierta trayectoria antes de caer al suelo, está claro que no ha trazado una línea recta, que sería el camino más

corto. Lo que se ha minimizado no ha sido la longitud, sino la acción, de modo que tenemos que explicar qué significa esto.

En cada uno de los puntos de su trayectoria, una partícula en movimiento tiene una posición x y una velocidad v, y en correspondencia, una energía cinética Ecinética = ½ mv^2 y una energía potencial V(x). Definimos el lagrangiano L(x, v), por el matemático francés Joseph-Louis Lagrange, como la energía cinética menos la energía potencial:

lagrangiano = energía cinética energía potencial

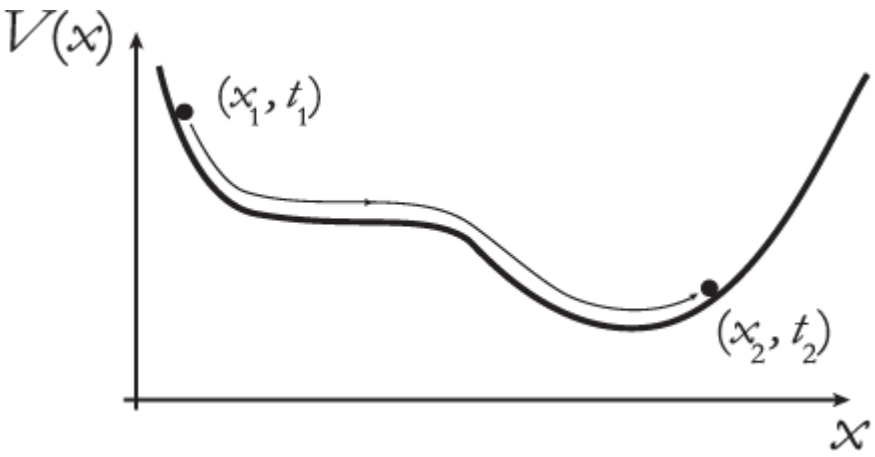
$$L(x, v) = \frac{1}{2}mv^2 - V(x) \qquad (3.14)$$

Así, dada una trayectoria [x(t), v(t)], la acción S de esa trayectoria es la integral del lagrangiano en un lapso de tiempo:

$$S = \int L(x, v)dt \qquad (3.15)$$

El lagrangiano tiene algún valor numérico en cada punto de la trayectoria, es decir, en cada momento. Pero la acción no es una función referida al punto de la trayectoria en que nos encontremos, sino que es una función de la trayectoria entendida en su totalidad. La acción de la trayectoria real que tomará la partícula, dadas sus posiciones y tiempos iniciales y finales, será más pequeña de lo que sería la acción para cualquier otra trayectoria que pudiéramos imaginar entre esos puntos. Este modo de formular la mecánica clásica se conoce a menudo como mecánica lagrangiana, ya que el lagrangiano es la cantidad central de interés. La acción es la integral del lagrangiano en el tiempo, y los movimientos físicos reales son aquellos que minimizan la acción.

Veamos cómo funciona el principio de mínima acción para una partícula en un potencial. Parte de la posición x1 en un tiempo t1, y termina en la posición x2 en el tiempo t2.



Puede que piense: «Dejaré ir la partícula en x1 y calcularé su aceleración mediante F = ma, luego iré integrando para obtener la velocidad, hasta que llegue a x2». Si es así, está todavía estancado en el modo de pensar laplaciano. No podemos dejar ir simplemente la partícula con velocidad inicial cero, porque podría alcanzar el punto final x2 a destiempo. En este mismo sentido, no es un requisito que la pelota se detenga en el tiempo t2, sino solo que esté localizada en x2 en ese momento. El hecho de que conozcamos los momentos inicial y final, así como tales posiciones, es una fuerte restricción con respecto a lo que pueda hacer entremedias la partícula. Lo que la partícula hará realmente será seguir la trayectoria que minimice su acción, sujeta a dichas restricciones, y este requisito nos dirá cuál habrá tenido que ser la velocidad inicial. Quizás haya tenido que remontar la cuesta de partida, alejándose de x2, para no alcanzar su destino final demasiado pronto. La velocidad terminará siendo la que haya tenido que ser para garantizar que la partícula termina donde se le supone en el momento señalado.

¿Qué significa minimizar la acción? La acción es la integral del lagrangiano, que es la energía cinética menos la energía potencial. La energía cinética ½ mv^2 nunca es un número negativo, de modo que para minimizar la acción queremos hacer que la energía cinética sea lo más

pequeña posible. Pero no podemos llevarla exactamente a cero, pues en tal caso la partícula no se movería, y por tanto no llegaría a x^2 a su debido tiempo. Tiene que moverse lo suficientemente rápido para recorrer la distancia adecuada en el tiempo oportuno, pero no demasiado rápido si puede evitarlo.

Por lo demás, minimizar «menos el potencial» parece bastante fácil: no hay más que desplazarse a algún punto en que V(x) sea realmente alto, de modo que −V(x) sea realmente bajo. (Para este propósito, un número negativo grande cuenta como «bajo»). Pero esto entra en conflicto con el párrafo anterior. Si partimos con la partícula en x1 y pasamos rápidamente a un punto en que el potencial es mucho mayor, tendrá que retroceder rápidamente para llegar a x2 en el momento oportuno, y eso aportará una energía cinética elevada. Estamos intentando mantener la energía cinética integrada lo más pequeña posible.

Minimizar la acción establece un acto equilibrador: la energía potencial tiende a ser alta, pero la energía cinética tiende a ser baja, y ambas tienen que llegar a un arreglo con el fin de comenzar y terminar en los puntos especificados y en los momentos especificados. Como resultado, la trayectoria que minimiza la acción resulta ser precisamente la que cumple las ecuaciones clásicas del movimiento. El modo mínima acción de hacer las cosas es matemáticamente equivalente a la formulación original de Newton.

Lo cual es notable, si se piensa. La segunda ley de Newton, el principio nuclear de la mecánica clásica, establece que la fuerza que actúa sobre un objeto es igual al producto de su masa por su aceleración. Desde la perspectiva de la mínima acción, la palabra «fuerza» ni siquiera aparece por ninguna parte. Al final se trata de la misma dinámica, pero para llegar a término ha entrado en juego un conjunto de conceptos totalmente distinto.

Así pues, ¿cuál es la perspectiva correcta? ¿Parte la naturaleza realmente de un estado inicial y va avanzando paso a paso, un instante tras otro, como nos daría a entender Laplace? ¿O tiene la naturaleza algún tipo de precognición, por la que puede visualizar todos los posibles movimientos que pudieran darse entre un punto inicial y un punto final, para elegir aquel que minimiza la acción?

Ni una ni otra. La naturaleza no es más que naturaleza, y hace lo que hace. Nosotros, los seres humanos, hacemos lo posible por entenderla con arreglo a nuestros propios términos. Puede resultar que descubramos diferentes formas equivalentes de conceptualizar un mismo comportamiento subyacente. En tales casos, no es tan importante preocuparse acerca de cuál es la «correcta», como estar preparados para pensar en cualesquiera términos que ofrezcan el mayor campo de percepción de la situación que nos ocupe.

1 Los tres últimos se tomaron de los nombres de sendas mascotas publicitarias creadas en los años treinta del siglo XX para una marca de cereales. (N. del T.)

2 ¿Puede ver por qué tiene que ser a «casi» todo? Si tuviéramos mala suerte, podríamos dar con un sistema en que c = d = 0, por ejemplo, cuando el potencial fuera precisamente $V(x) = ex^4$. En tal caso no existe una aproximación en que el oscilador sea armónico, ni siquiera para una x pequeña.

CUATRO

ESPACIO

Hasta ahora hemos estado hablando de cosas que suceden: pelotas que ruedan por pendientes, pesos que oscilan del extremo de un muelle, planetas que giran en torno al sol… En todo ello está implícita la idea de un escenario en el que tales cosas tienen lugar. Ese escenario es el espacio, el conjunto de todas las localizaciones posibles de las cosas. Nos hemos referido a la idea de espacio repetidamente —ahí estaba, en el primer párrafo del Capítulo 1— sin demasiada deliberación, puesto que presumiblemente usted está en cierta medida familiarizado con el concepto. Ha llegado el momento de profundizar un poco más. ¿Qué propiedades tiene el espacio, y por qué? ¿Por qué necesitamos de algo llamado espacio, para empezar?

Puede que algunas de estas preguntas tengan una respuesta satisfactoria; puede que otras no. Lo que podemos hacer es pensar acerca de las propiedades que tiene el espacio, y cómo se relacionan con otras características del mundo físico. El espacio, ¿es algo? ¿O es tan solo una propiedad de las cosas? ¿Qué significa decir que el espacio es tridimensional? Y ¿cuáles son los comportamientos físicos que hacen que sea útil pensar en términos de espacio? Estas preguntas nos conducirán a investigar otra formulación más de la mecánica clásica: después de la mecánica newtoniana y de la mecánica lagrangiana, examinaremos una perspectiva llamada mecánica hamiltoniana. A diferencia de las otras formulaciones, la mecánica hamiltoniana no trata de partida al espacio como poseedor de un derecho especial. Es, por tanto, un contexto útil en el que pensar acerca de las propiedades especiales del espacio.

SUSTANCIA FRENTE A RELACIÓN

A comienzos del siglo XVIII, en los albores de la mecánica clásica, se pensó mucho acerca de la cuestión de lo que el espacio «es realmente». Una idea era que el espacio fuera una sustancia por sí mismo, con una existencia separada de las otras cosas contenidas en él. El mundo, según este punto de vista, consiste en toda una serie de objetos, más el espacio en el que están engastados. El espacio actúa como un contenedor para todo lo que está dentro de él.

Esto nos suena bastante natural, pero hay una perspectiva alternativa: que el espacio no sea nada, sino meramente una forma de reformular el hecho de que entre dos objetos cualesquiera se da una característica cuantificable a la que llamamos «distancia entre ellos». Desde este punto de vista, el espacio es fundamentalmente relacional. Una vez que sabemos cómo están relacionados todos los objetos del mundo por las distancias que los separan, no existe ninguna cosa extra llamada «espacio» en la que dichos objetos residan. Cualquier mención al «espacio» no es más que una forma conveniente para describir tales relaciones de distancia.

Estas dos perspectivas opuestas acerca de la naturaleza del espacio, «sustancialismo» y «relacionalismo», fueron objeto de un animado intercambio epistolar entre Samuel Clarke desde Inglaterra y Gottfried Wilhelm Leibniz desde Alemania. Recordará a Leibniz como el codescubridor del cálculo junto con (y feroz rival de) Isaac Newton. Ambos tuvieron algún que otro fogoso intercambio, pero Newton prefería trabajar a través de intermediarios como Clarke.

La correspondencia entre Leibniz y Clarke fue instigada por Carolina de Ansbach, una noble educada en Prusia, donde Leibniz era uno de sus tutores. Se trasladó a Inglaterra cuando su esposo, George Augustus, fue nombrado Príncipe de Gales; más tarde se convertiría en reina, cuando él fue coronado como el rey Jorge II. Carolina sentía un duradero interés por la ciencia y la filosofía. Se convirtió en paladín de la variolización, precursora de la moderna vacunación, después de dirigir un estudio sobre su adecuación para combatir la viruela.

Después de que Carolina llegara a la corte británica para convertirse en princesa, Leibniz le escribió para prevenirla contra ciertos aspectos de la filosofía inglesa que él consideraba teológicamente problemáticos, con especial referencia a John Locke y por supuesto a Isaac Newton. Carolina —¿con travesura?— llamó la atención de Clarke acerca de tales críticas. Clarke, destacado clérigo anglicano y amigo de Newton, escribió a Leibniz para defender el punto de vista de Newton según el cual el espacio era absoluto, frente a la perspectiva relacional de Leibniz, fuertemente influenciada por Descartes. Sabemos que Newton, como mínimo, ofreció consejo a Clarke sobre lo que debía escribir, si bien algunos estudiosos sostienen que las cartas habían sido escritas secretamente por Newton.

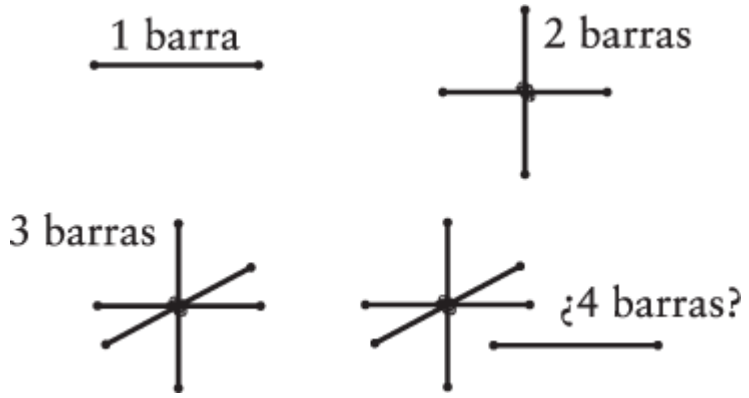
La correspondencia entre Leibniz y Clarke quedó bruscamente interrumpida por la muerte de Leibniz en 1716, pero las cartas que han llegado hasta nosotros representan una obra mayor en torno a la temprana filosofía de la ciencia. Tratan no solo de la naturaleza del espacio, sino también del libre albedrío y la esencia de Dios. Hoy, la mayor parte de los físicos se pondrían del lado de Newton, considerando el espacio como algo con existencia propia, por un par de razones. En primer lugar, el espacio entre dos objetos no está vacío; está ocupado por campos de varios tipos. En segundo lugar, el espacio (como parte del espacio-tiempo), tiene vida propia; Einstein mostró que la geometría del espacio responde a la energía y puede cambiar con el tiempo, como veremos en el Capítulo 8. Pero no tenemos la respuesta final.

Y así, una cuestión que durante un tiempo se consideró absolutamente central para el progreso de la física ha terminado archivada en un cajón, sin haber obtenido una respuesta definitiva. Parte de la evolución de la ciencia no consiste simplemente en el progreso asociado al aprendizaje de cosas nuevas; está también el proceso por el que decidimos qué preguntas son importantes y cuáles pueden ser ignoradas. A veces es porque la pregunta no es verdaderamente interesante; otras veces simplemente no es el momento adecuado para una respuesta. De acuerdo con una estimulante idea moderna, el espacio surgiría del entrelazamiento cuántico, en lugar de constituir una sustancia de pleno derecho, de modo quizás en algún sentido pueda decirse que el relacionalismo vaya a triunfar al final.

DIMENSIONES

Dando por sentado que poseemos una comprensión más o menos intuitiva de lo que se quiere significar con «espacio», pasemos a considerar algunas de sus propiedades. La propiedad más importante es su dimensionalidad. El espacio por el que nos movemos en nuestra vida cotidiana es tridimensional. Pero los físicos con frecuencia apuntan a espacios de dimensiones menores o mayores. En ocasiones porque consideran una o dos dimensiones espaciales a la hora de realizar modelos de juguete simplificados; otras veces, al postular más de tres dimensiones espaciales en teoría de cuerdas o en otros modelos de unificación. También existen otros «espacios» matemáticos abstractos —conjuntos con alguna estructura adicional, como el espacio de configuración o espacio de fases de un sistema de muchas partículas—, que pueden tener un número completamente diferente de dimensiones. En este capítulo nos centraremos en el entrañable espacio de siempre que nos rodea.

Imaginemos que cogemos dos barras largas y delgadas, y que las sujetamos de modo que queden perpendiculares entre sí (formando ángulos rectos). Cogemos ahora una tercera barra y la sujetamos a las dos primeras, en el mismo punto de intersección, de tal forma que esta nueva barra sea perpendicular a las dos de inicio. Coja una cuarta barra y sujétela en el mismo lugar, de manera que sea perpendicular a las tres que ya teníamos.

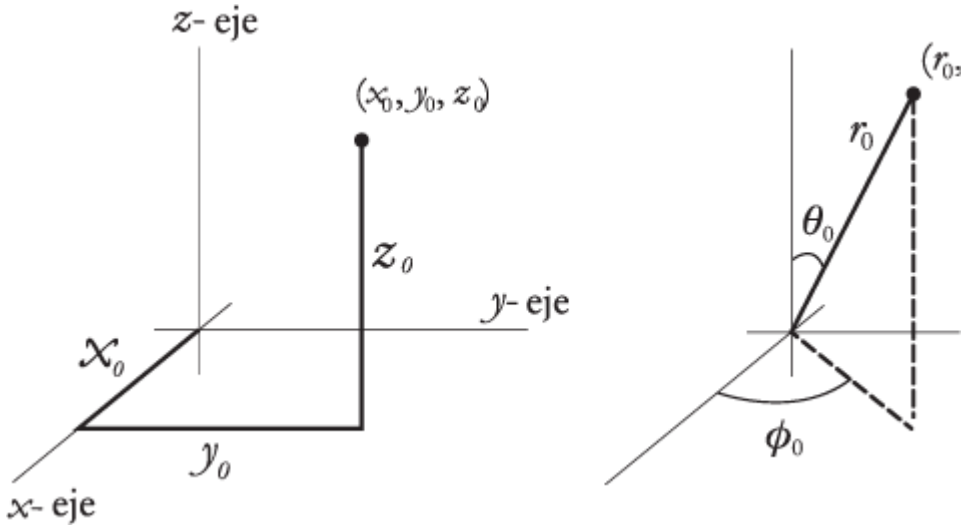


No puede. En nuestro mundo, tres líneas pueden ser mutuamente perpendiculares entre sí, pero cuatro ya no. Esta es una forma de demostrar que el espacio es tridimensional.

(Seguramente es consciente de que se necesitan tres patas para que un taburete sea estable. ¿Cuántas patas se necesitarían en un espacio de dos dimensiones? ¿O de cuatro dimensiones, o de cualquier otro número?).

Una dimensión, entonces, es en esencia «una dirección independiente en que pueden moverse las cosas». Con esta consideración, moverse en un sentido o en el opuesto cuenta como la misma dirección. Determinándolo a partir de su propia posición actual, adelante/atrás sería una dirección, derecha/izquierda sería otra y arriba/abajo sería la tercera. Todas estas direcciones son independientes, mientras que podrían pensarse otras direcciones como combinaciones de estas. Tres dimensiones.

Otra forma de expresar esta misma idea es constatando que se necesitan tres números, llamados coordenadas, para especificar la localización de un punto en el espacio. Imagine que tomamos un punto de partida cualquiera y trazamos tres direcciones perpendiculares abstractas, o ejes, a modo de versión imaginaria de nuestras tres barras perpendiculares. Podemos asignar coordenadas a cualquier punto —por ejemplo, (x, y, z)— yendo del origen a ese punto mediante una serie de segmentos rectos perpendiculares, y midiendo luego la longitud de cada segmento.



Lo importante no es el sistema concreto de coordenadas que usemos, sino el hecho de que se requerirán siempre tres números para especificar la localización de un punto. El conjunto de ejes perpendiculares define las coordenadas cartesianas, llamadas así por René Descartes. Pero podríamos igualmente especificar el mismo punto midiendo r, la distancia de un segmento radial desde el origen al punto; (la letra griega «zeta»), el ángulo polar entre ese segmento radial y el eje z; y ϕ(la letra griega «fi»), el ángulo acimut entre el eje x y la proyección de nuestro punto hacia abajo, al plano xy. Se llama a estas coordenadas esféricas, dado que el conjunto de todos los puntos para cierto valor dado de r define una esfera.

Las coordenadas no son atributos objetivos del mundo. Son invenciones humanas, etiquetas que ponemos a diferentes localizaciones en el espacio. La regla general es que no hay ninguna cantidad física que dependa de las coordenadas que usemos, de la misma manera que la longitud física de un objeto no depende de si viene expresada en centímetros o en pulgadas. Esta propiedad de la invariancia de las coordenadas parece bastante claro, pero otras versiones retocadas de la misma nos llevan a contemplar simetrías y teorías de campo de gauge, que desempeñan un papel central en la física fundamental moderna.

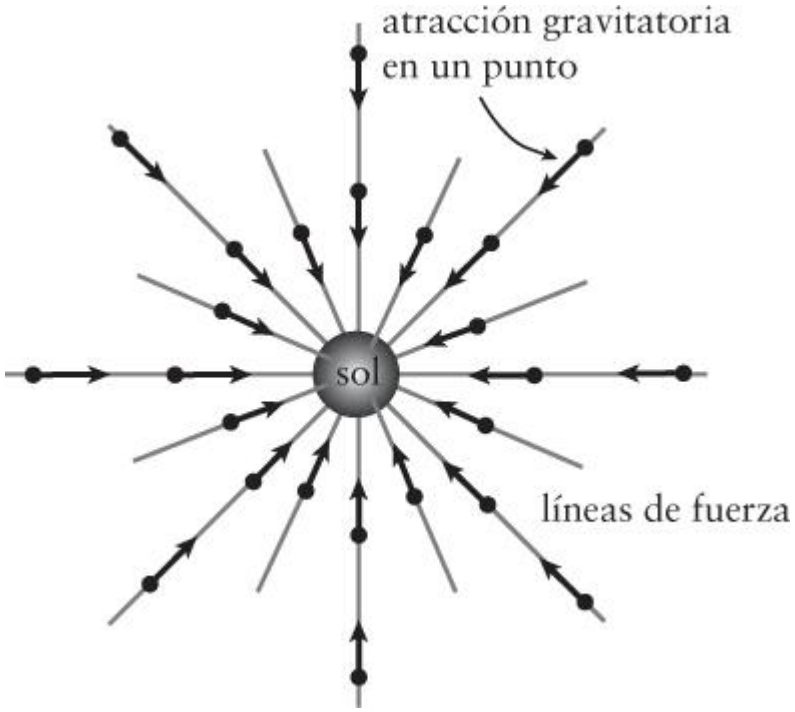
Si el espacio es tridimensional, podría haber subconjuntos de espacio que fueran de una dimensión menor, al menos como buena aproximación. Considere un hilo conductor largo y recto. Realmente es tridimensional, por

supuesto: si lo observamos de cerca, podemos ver que la sección del hilo no tiene valor cero. Pero si lo miramos de lejos, la apariencia que nos muestra es en esencia unidimensional. Si los electrones que se desplazan a lo largo del hilo tienden a hacerlo rápida y libremente en toda su longitud, y no siguiendo direcciones perpendiculares (o muy difícilmente), entonces los físicos modelarán con gran gozo ese sistema como si fuera de una sola dimensión. De forma similar, hay casos en que ciertos fenómenos interesantes son confinados a espacios bidimensionales, tales como películas muy delgadas o superficies de objetos tridimensionales; en estos casos, a menudo es posible tratar provechosamente los fenómenos relevantes como si realmente se desenvolvieran en un mundo de dos dimensiones. De nuevo la filosofía de la vaca esférica en acción: realizar una aproximación de un sistema complicado a otro más simple, tanto que desaparezcan dimensiones enteras.

DIMENSIONES Y FUERZAS

La vida sería completamente distinta —de hecho la vida biológica probablemente no fuera siquiera posible— si el espacio tuviera un número diferente de dimensiones. Consideremos una de nuestras fuerzas físicas favoritas, la gravedad, descrita por la ley de la inversa del cuadrado de Newton: la fuerza gravitatoria entre dos objetos es inversamente proporcional al cuadrado de la distancia entre ellos. ¿Por qué es la inversa del cuadrado de la distancia? ¿Por qué no la distancia misma, o la distancia elevada a la decimoctava potencia?

Una simple visualización podrá resultarnos útil. Visualice mentalmente una multitud de líneas rectas que parten del sol hacia el infinito en todas direcciones. Son líneas de fuerza, llamadas así porque la fuerza de gravedad atrae a los objetos a lo largo de ellas. Son puramente conceptuales, no existen en la realidad líneas físicas de ningún tipo que partan del sol. Pero visualizar esas líneas diseminándose en todas direcciones nos ayuda a explicar la ley de la inversa del cuadrado.



Si dibujamos una esfera circunscrita al sol, todas las líneas de fuerza pasan a través de esa esfera. Si dibujamos otra esfera con un radio mayor, la atravesarán las mismas líneas, pero estarán más separadas: pasarán menos líneas a través de cualquier área fija de la esfera.

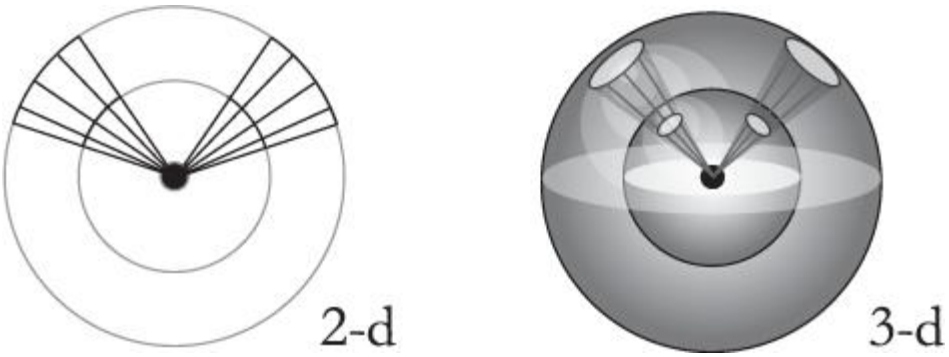
El área total A de una esfera guarda una relación con su radio r expresada como $A = 4\pi r^2$. La expresión 4π de la fórmula es algo que solemos memorizar sin más, pero el hecho de que el área sea proporcional a r^2 es algo que se obtiene por análisis dimensional. Las cantidades físicas se expresan siempre en términos de unidades: longitud, tiempo, masa, o alguna combinación de varias.[1] Siempre que sumamos cantidades, o las relacionamos mediante ecuaciones, tienen que venir expresadas en las mismas unidades.

Las unidades del área se expresan siempre en términos de «cuadrado de la distancia» (porque se trata de una cantidad bidimensional), y la única distancia implicada en la cuestión que nos ocupa (¿cuál es el área de una esfera de radio r?) es el radio. Así que, naturalmente, el área será proporcional a r^2; no podría ser de otro modo. De forma similar, su

circunferencia se mide en unidades de distancia, por lo que será proporcional a r (es $2\pi r$).

Las líneas de fuerza gravitatorias no comienzan ni terminan en un espacio vacío; se originan en objetos con una masa, que son la fuente de la gravedad. Así, cuando consideramos las líneas que emanan del sol, el número total de líneas que pasan por cada esfera es el mismo, y los radios de las esferas son diferentes. El área de esas esferas aumenta en proporción a r^2. Como resultado, unas líneas que estaban densamente agrupadas cerca del sol, se separan a medida que nos alejamos. En concreto, la densidad de las líneas decrece a razón de $1/r^2$. Y el número de líneas que pasa a través de un objeto es proporcional a la intensidad de la fuerza gravitatoria.

Y ahí tenemos la ley de la inversa del cuadrado. La fuerza experimentada por un objeto simplemente es proporcional a la densidad de líneas de fuerza que pasan cerca de él, que decrece de manera inversamente proporcional al cuadrado de la distancia, porque las líneas se dispersan por la superficie de una esfera cuya área crece en proporción con el cuadrado de la distancia. (Esta misma lógica explica por qué los objetos parecen más apagados según nos alejamos de ellos: simplemente tenemos que remplazar «líneas de fuerza» por «rayos de luz»).



Las cosas serían muy diferentes si el espacio tuviera un número diferente de dimensiones. En un espacio bidimensional, un objeto estaría alojado en un círculo unidimensional, en lugar de estarlo en una esfera bidimensional. Entonces, la densidad de las líneas de fuerza, y por tanto la intensidad de la atracción de la fuerza gravitatoria, disminuiría en proporción a la inversa de

la distancia, y no del cuadrado de la distancia. Por otra parte, en un espacio de cuatro dimensiones un objeto estaría alojada en una hiperesfera de tres dimensiones. (Llegados a este punto es donde resulta muy difícil visualizar las cosas, por lo que tenemos que confiar directamente en las matemáticas). Así, la gravedad obedecería una ley de la inversa del cubo. En general, en un universo en que el espacio es de un número d de dimensiones, la fuerza de la gravedad es proporcional a 1/ rd−1.

Vistas así las cosas, parece casi inevitable que las fuerzas de la naturaleza obedezcan a leyes de la inversa del cuadrado. Pero no es así… al menos no siempre. Al nivel de las partículas elementales, existe una «fuerza nuclear fuerte» y una «fuerza nuclear débil» que actúan sobre radios de acción muy cortos antes de disminuir rápidamente hasta prácticamente desaparecer. Las razones son diferentes en cada caso. En el de la fuerza nuclear fuerte, las líneas de fuerza se enredan unas con otras en lugar de diseminarse hacia el infinito; en el caso de la fuerza nuclear débil, las líneas parecen atenuarse gradualmente, pero en realidad es porque son absorbidas por un campo de Higgs que se extiende por todo el espacio. Nadie dijo nunca que la naturaleza fuera ordenada.

Hay un ejemplo famoso, al margen de la gravedad, de ley de la inversa del cuadrado: el electromagnetismo. El campo eléctrico de una partícula con carga eléctrica genera líneas de fuerza que se dispersan al infinito (a menos que choquen con toras partículas con carga eléctrica, cosa que sucede a menudo). Como resultado, la intensidad de la fuerza eléctrica obedece a una ley de inversa del cuadrado, conocida en este caso como ley de Coulomb. Al menos en el espacio tridimensional.

La idea de las líneas de fuerza fue sugerida por Michael Faraday a mediados del siglo XIX, dentro del ámbito del electromagnetismo. Faraday había nacido en el seno de una familia pobre, era hijo de un herrero rural. Al llegar a la adolescencia, se empleó como aprendiz en el negocio de un librero local, circunstancia que lo proporcionó la oportunidad de educarse a través de la lectura. Entró en la Royal Institution de Londres, trabajando primero como químico (fue el descubridor del benceno y el inventor de una versión primitiva del quemador Bunsen), antes de caer fascinado por la electricidad y el magnetismo. Su trabajo en este campo preparó el terreno

para que luego James Clerk Maxwell reuniera estos fenómenos en una teoría unificada del electromagnetismo. Maxwell poseía el bagaje matemático necesario para codificar las intuiciones físicas de Faraday bajo la forma de un sistema de rigurosas ecuaciones.

EL MOMENTO Y LA VELOCIDAD REPENSADOS

¿Qué es lo que hace que el espacio sea «espacio»? Es decir, ¿cuáles son las propiedades que tiene el mundo y que nos mueven a describirlo como «un conjunto de cosas distribuidas por el espacio»? (Y que evolucionan en el tiempo, pero esto lo dejamos para el capítulo siguiente).

Para abordar esta cuestión, hagamos una de esas cosas que tanto nos gustan: introducir una nueva forma de pensamiento en torno a la mecánica clásica. Tenemos ya la forma newtoniana de elaborar la mecánica clásica, consistente en especificar las posiciones y velocidades de ciertas partes de un sistema, y utilizar las leyes newtonianas para determinar su evolución en el tiempo. Y tenemos la mecánica lagrangiana, con su principio de mínima acción, que nos induce a imaginar todas las trayectorias que puede seguir un sistema entre unas condiciones específicas, inicial y final, y elegir la trayectoria con mejor acción como aquella que el sistema sigue realmente.

Nuestro tercer modo de formular la mecánica clásica es la mecánica hamiltoniana. Su idea central consiste en erigir el concepto de «momento» como algo con existencia propia, independiente de la «velocidad». Este paso puede ser un poco delicado, porque a primera vista parece casi indistinguible del formalismo newtoniano, pero es diferente (y más potente) en un sentido que puede resultar sutil. Pero tales sutilezas serán cruciales para entender por qué el espacio es un concepto tan importante.

En el capítulo anterior hablábamos acerca del espacio de fases, el conjunto de todas las posibles posiciones y momentos que podía tener un sistema. De acuerdo con el paradigma laplaciano, especificar un punto en el espacio de fases en un instante del tiempo es suficiente para determinar la trayectoria

entera de un sistema (al menos de un sistema que esté protegido de influencias exteriores). Y aunque utilicemos el momento como parte del espacio de fases, sabemos que en el marco de la mecánica newtoniana, momento y velocidad son intercambiables, por cuanto se relacionan de la forma .

Por feliz que pueda parecernos esta imagen, hay un pequeñísimo detalle que resulta extraño, tan pequeño que lo pasamos por alto en su momento. Se trata de lo siguiente: lo que con tanto éxito postula el paradigma laplaciano es que si nos dicen cuál es el estado del sistema en un simple instante del tiempo, podemos determinar el resto de la trayectoria. Pero, ¿realmente «la velocidad» es algo susceptible de determinarse en un simple instante del tiempo? La velocidad es la derivada de la posición: la razón de cambio de la posición con respecto al tiempo. Para calcular esta derivada, necesitamos mirar más allá de cualquier simple instante, preguntarnos qué está haciendo el sistema un instante después. Aunque ese instante sea un después infinitesimal con respecto al instante original, sigue pareciéndonos que habremos tenido que asomarnos mínimamente a otros instantes en el tiempo, lo cual encaja pésimamente con la filosofía laplaciana.

LA MECÁNICA HAMILTONIANA

La perspectiva hamiltoniana resuelve este desajuste de un modo elegante. Una vez más, partimos del espacio de fases, considerado como el conjunto de todas las posiciones y momentos que puede tener un sistema. Pero ahora destacaremos que es en realidad el vector momento de lo que estamos hablando, no de la velocidad. De hecho, desde el punto de vista hamiltoniano, no definimos el momento como el producto de la masa por la velocidad. Tomamos el momento como un concepto completamente independiente, en condiciones de igualdad con respecto a la posición, consideradas ambas las dos cosas que necesitamos para especificar un punto fijo en el espacio de fases. Una partícula (o un sistema más complicado) está localizada en un punto del espacio y se da además en ella propiedad de tener un vector, el «momento». Esto es así para cualquier

simple instante en el tiempo, ya que de ahora en adelante no pensamos ya en el momento como algo relacionado con la razón de cambio de nada.

Así es como funciona la dinámica hamiltoniana. Partimos del espacio de fases, el conjunto de todas las coordenadas x y momentos p. (Para mayor simplicidad, no nos preocuparemos por utilizar las flechitas sobre las letras para señalar que se trata de vectores, ni subíndices para indicar de qué parte del sistema estamos hablando; vamos a imaginar que todo ello está implícito). Luego definimos una función, llamada hamiltoniano: H(x, p). El hamiltoniano es básicamente la energía del sistema, expresada en términos de posiciones y momentos.

Sabemos que la energía potencial del sistema depende únicamente de las posiciones, por lo que la expresamos como V(x). Tenemos además la energía cinética, que en la mecánica newtoniana expresábamos como $K = \frac{1}{2}mv^2$. Pero nos gustaría expresar esto en términos del momento, no de la velocidad. En la mecánica newtoniana, el momento y la velocidad se relacionaban: $p = mv$. Así, podemos obtener una expresión para la energía cinética en términos del momento sustituyendo $v = p/m$ para obtener $K = p^2/(2m)$. Recapitulando, por tanto, podemos escribir:

hamiltoniano = energía cinética + energía potencial

$$H(x,p) = \frac{p^2}{2m} + V(x) \qquad (4.1)$$

Hasta ahora, no hemos hecho más que algunas manipulaciones para rescribir la energía en términos de posición y momento (y no de velocidad). La estupenda novedad de esto es que podemos derivar las ecuaciones de movimiento para el sistema a partir de esta única expresión. Explicaremos cómo hacerlo enseguida, pero antes que nada vamos a saltar directamente a las respuestas para darnos una pequeña recompensa.

En la mecánica newtoniana, seguimos la evolución de tan solo una variable, la posición x(t) como una función de tiempo. Cosas como la velocidad y la aceleración se definen en relación con esta variable (respectivamente como primera y segunda derivada). Correspondientemente, solo se necesita una ecuación para explicar lo que sucede: F = ma. Pero en la mecánica hamiltoniana hay dos variables, x(t) y p(t). Así que ahora necesitamos dos ecuaciones. Y de hecho las tenemos: las dos ecuaciones para las derivadas del momento y de la posición, dadas por

$$\frac{dp}{dt} = -\frac{dV}{dx} \qquad (4.2)$$

y

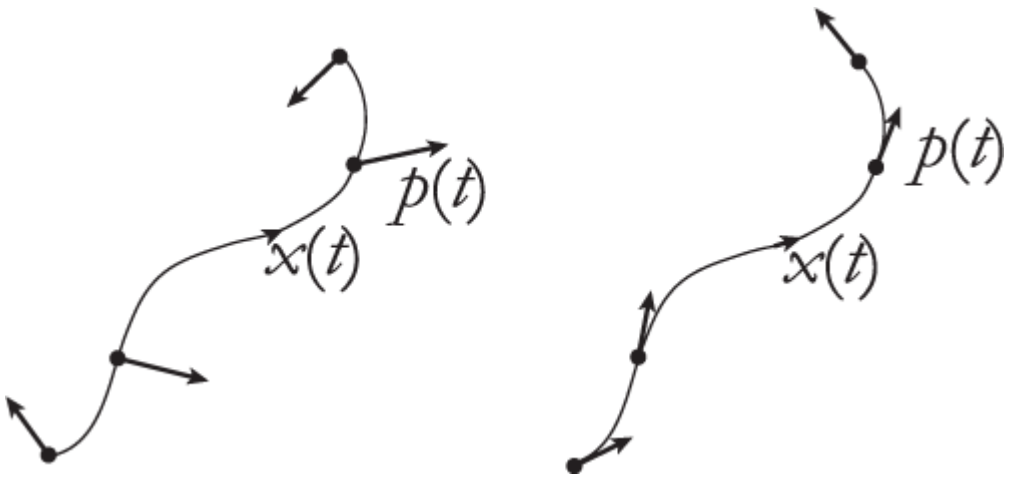
$$\frac{dx}{dt} = \frac{p}{m} \qquad (4.3)$$

La primera de estas dos ecuaciones nos resulta familiar, es simplemente la segunda ley de Newton, F = ma, con diferente notación. Recuerde, por (3.13), que podemos remplazar ma por dp/dt, y por (3.3) sabemos que la fuerza es menos la derivada del potencial con respecto a la posición. La segunda ecuación, (4.3), nos resulta igualmente familiar: no es más que una reformulación de p = mv, el momento es el producto de la masa por la velocidad. Pero la interpretación de esta ecuación es diferente, y la diferencia es crucial. Captar la distinción es comprender la mecánica hamiltoniana hasta la médula.

Según la perspectiva newtoniana, el objeto fundamental es x(t), la trayectoria definida por la posición como una función de tiempo. A partir de ahí todo se sigue. Podemos elaborar una derivada para obtener la velocidad, y definir el momento como el producto de la masa por la velocidad. Pero el momento no se define de este modo en la mecánica hamiltoniana: es una variable de pleno derecho. La relación (4.3) parece la misma en ambas

perspectivas, pero desde el punto de vista hamiltoniano es una ecuación de movimiento, una relación que se mantiene verdadera en las trayectorias reales del mundo físico, no algo que es verdad por definición.

¿Qué diferencia hay entre una definición y una ecuación de movimiento? Una definición implica inevitabilidad en la relación; en la mecánica newtoniana, no hay nada que el momento pudiera ser más que el producto de la masa por la velocidad, puesto que así se ha definido. Una ecuación de movimiento, por el contrario, elige la relación «correcta» entre diversas variables, pero siempre teníamos la alternativa de elegir entre otras variables (incluso erróneas).



Dos trayectorias en un espacio de fases, en las que suponemos un punto en el espacio x(t) y un momento p(t) para cada tiempo t. A la izquierda, los vectores momento no son proporcionales a la velocidad —no pueden serlo, puesto que no apuntan en la dirección del camino espacial—, de modo que la trayectoria no obedece a las ecuaciones de movimiento, mientras que la de la derecha sí lo hace.

En la mecánica hamiltoniana, el momento tiene una existencia independiente de la velocidad. Somos libres para imaginar cualesquiera trayectorias para la posición y el momento, x(t) y p(t), incluidas las hipotéticas en que p no fuera igual a mv. En el espacio de todos los comportamientos concebibles, no existe una relación necesaria entre

posición y momento. Pero hay ciertas trayectorias especiales, aquellas que obedecen a las ecuaciones de movimiento. Y en esas trayectorias, el momento es igual al producto de la masa por la velocidad.

Si su cerebro insiste en concebir el momento como el producto de la masa por la velocidad, imagine que en la discusión anterior sustituye «momento» por cualquier otra etiqueta, «purpureidad» por ejemplo. Así, la mecánica hamiltoniana dice que el estado de un sistema en un instante del tiempo cualquiera viene definido por su posición y su vector purpureidad. En aquellas trayectorias físicas que obedecen a las ecuaciones de movimiento, sucede que la purpureidad es igual al producto de la masa por la velocidad, pero podemos imaginar otras en que esto no se da. En cierto modo nos resulta poco conveniente utilizar la etiqueta «momento» tanto para la variable independiente en el espacio de fases en la imagen hamiltoniana, como para el producto de la masa por la velocidad en la imagen newtoniana; ambas son numéricamente iguales cuando una trayectoria obedece a la ecuación de movimiento, pero son conceptualmente distintas.

ECUACIONES DE HAMILTON

Esta es la filosofía de la mecánica de Hamilton: tratar la posición y el momento como dos variables conceptualmente distintas, y determinar ecuaciones de movimiento para cada una de ellas. El «hamiltoniano» en sí mismo es simplemente la energía del sistema expresada como una función de posición y momento (y no como una función de velocidad u otras derivadas). La idea es que deberíamos ser capaces de utilizar la expresión explícita del hamiltoniano, (4.1), para derivar ecuaciones de movimiento para el momento (4.2) y la posición (4.3). ¿Cómo funciona esto exactamente?

Pensemos en la ecuación de movimiento para el momento, (4.2). Dice que la razón de cambio del momento en el tiempo es igual a (menos) la pendiente de la función potencial-energía. Esto es moralmente equivalente a la segunda ley de Newton, F = ma, puesto que la pendiente del potencial es

(menos) la fuerza ejercida sobre el objeto. Podemos pensar que el momento es zarandeado a expensas de los cambios en la energía potencial debidos a la posición. Tiene sentido, desde el punto de vista de la física: si el potencial fuera completamente plano (pendiente cero, como una pelota rodando sobre una mesa lisa), el momento se conservaría.

Pero la energía potencial, advertimos, es uno de los dos términos del hamiltoniano. El otro es la energía cinética, $p^2/2m$. Si el momento estaba a expensas de la pendiente de la energía potencial, ¿es posible que se dé una especie de bonita simetría, y que la posición se vea alterada por la pendiente de la energía cinética?

Eso es exactamente lo que sucede. Recordemos la ecuación (2.7): la derivada de x^2 con respecto a x es 2x. Esto funciona sea cual sea la variable a considerar; así por ejemplo, la derivada de p^2 con respecto a p es 2p. La derivada de la energía cinética con respecto al momento es, por tanto:

$$\frac{d}{dp}\left(\frac{p^2}{2m}\right)=\frac{1}{2m}\frac{d}{dp}\left(p^2\right)=\frac{1}{2m}(2p)=\frac{p}{m} \qquad (4.4)$$

(Siempre que consideramos la derivada de una función, las constantes como 2m son meros acompañantes; en este caso, hay dos factores de 2 que se cancelan mutuamente).

¡Interesante! Tal como habíamos conjeturado, al tomar la derivada de la energía cinética con respecto al momento, obtenemos el término derecho de (4.3), la ecuación de movimiento para la posición. Hay un molesto digno menos en una de las ecuaciones y no en la otra, pero, en términos generales, el momento se ve alterado por cambios en la energía potencial, y la posición se ve alterada por cambios en la energía cinética. Resumiendo:

razón de cambio del momento en el tiempo = menos la pendiente de la energía potencial con respecto a la posición; razón de cambio de la posición

en el tiempo = pendiente de la energía cinética con respecto al momento

Juntas, ambas relaciones se conocen como ecuaciones de Hamilton.

Una bonita estructura, pero hay un punto técnico en el que debemos pensar un poco más. Para concretar las ideas, hemos estado centrándonos en una partícula simple moviéndose en un potencial unidimensional, un modelo de juguete estándar e ilustrativo. En este caso, el hamiltoniano resulta de la suma de las energías cinética y potencial, como en (4.1).

Pero el formalismo de la mecánica hamiltoniana es mucho más general. El hamiltoniano de un sistema siempre será una función de cierto conjunto de coordenadas y sus correspondientes momentos, pero hay una gran libertad con respecto a lo que pueda ser la función real. Podemos tener cualquier número de posiciones y momentos; en teoría de campos habrá un número infinito. De hecho, podemos considerar hamiltonianos muy complicados, algunos incluso en los que las posiciones y momentos se hallen mezclados, de forma que no sea posible distinguir entre «energía potencial» y «energía cinética». Gran parte de la física moderna consiste en conjeturar el hamiltoniano correcto para un sistema objeto de interés; una vez obtenido, sabemos cómo se comportará el sistema. Tenemos que pensar en términos un poco más generales.

DERIVADAS PARCIALES

Lo que buscamos es una versión de ecuaciones hamiltonianas que funcione para cualquier hamiltoniano. Y para obtenerla, vamos a tener que introducir por última vez un poco más de cálculo: las derivadas parciales.

Necesitamos esta tecnología porque queremos derivar nuestras ecuaciones de movimiento a partir de una sola expresión, el hamiltoniano total, en lugar de recurrir a la energía potencial y la energía cinética por separado.

Pero, a diferencia de la energía potencial, que depende únicamente de x, o de la energía cinética, que depende tan solo de p, el hamiltoniano depende de ambas, x y p, al mismo tiempo. Tenemos que ser capaces de obtener la derivada de una función de más de una variable.

La derivada de una función no es más que su pendiente para cada punto de la misma. Pero cuando la función depende de dos o más variables, no tenemos simplemente una curva con una única pendiente para cada punto. Es más probable que su paisaje represente un perfil serpenteante, en el que haya más de una dirección por la que podamos movernos. Algunas direcciones nos llevarán hacia arriba, otras hacia abajo, otras permanecerán en la misma altura. Deberemos tener en cuenta más matices, sin que sea posible definir «la pendiente» de una vez por todas.

Las derivadas parciales proporcionan una propuesta inteligente a esta cuestión. Consideremos una función simple de dos variables, x e y, bajo la forma del producto de x por y^2 por una constante a:

$$f(x, y) = axy^2 \qquad (4.5)$$

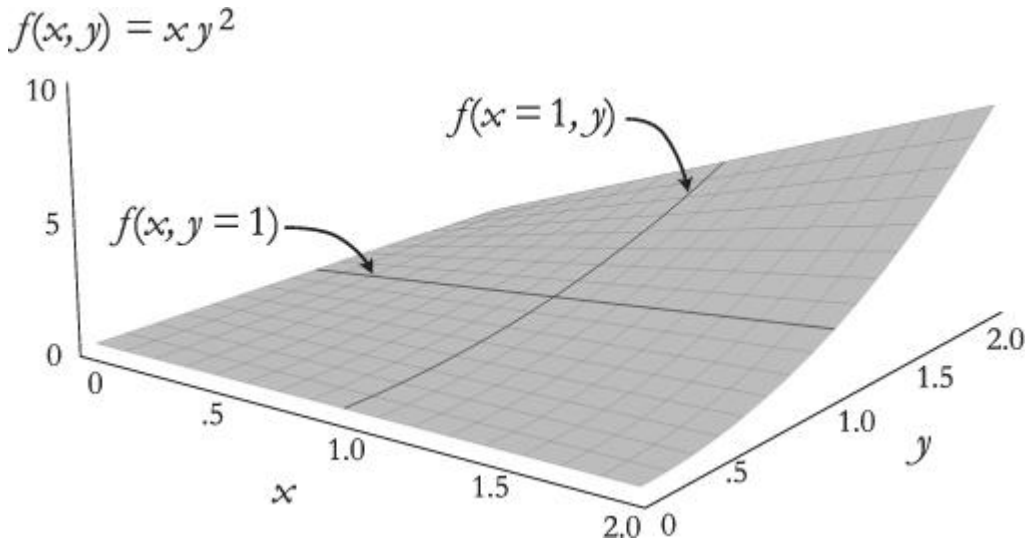
La idea que está detrás de las derivadas parciales es que cuando tenemos una función con varias variables, podemos abordarlas por turno. Es decir, obtenemos derivadas con respecto a cada variable por separado y, una vez hecho esto, tratamos todas las demás variables como si fueran constantes. Y para que los demás sepan lo que estamos haciendo, sustituimos la d de nuestras derivadas por un nuevo símbolo: ∂. (A veces lo llamamos «del», pero como esta palabra se usa para otros símbolos, es mejor llamarlo simplemente «parcial» cuando estemos leyendo en voz alta).

Por tanto, son dos las derivadas parciales que podemos obtener de f (x, y): la parcial con respecto a x y la parcial con respecto a y. Cuando obtenemos la parcial con respecto a x, diferenciamos con respecto a x, mientras que tratamos tanto a a como a y como constantes. Esto nos da:

$$\frac{\partial}{\partial x}\left(axy^2\right) = ay^2 \frac{d}{dx}(x) = ay^2 \qquad (4.6)$$

Mientras que cuando obtenemos la parcial con respecto a y, tratamos a a y a x como constantes:

$$\frac{\partial}{\partial y}\left(axy^2\right) = ax \frac{d}{dy}\left(y^2\right) = 2axy \qquad (4.7)$$



Una función de dos variables, f (x, y) = xy². Podemos obtener una derivada parcial con respecto a x pensando en y como una constante y diferenciando la función resultante; del mismo modo, podemos obtener una derivada parcial con respecto a y pensando en x como una constante.

Básicamente es esto. Para obtener una derivada parcial, especificamos cuál es la variable con respecto a la cual vamos a diferenciar y tratamos a todas las demás como si fueran constantes (a las que no afectará, por tanto, la operación de derivación). La tarea de resolver realmente ecuaciones diferenciales parciales puede resultar intricadamente difícil —hay personas

que consagran su vida profesional entera a encontrar mejores formas de hacerlo—, pero eso no nos preocupa mucho, aquí y ahora. Únicamente pretendemos ser capaces de ver de dónde proceden las ecuaciones de movimiento de Hamilton.

Podemos ahora recapitular. Sabemos que podemos elaborar ecuaciones de movimiento para el momento y la posición obteniendo la derivada de la energía potencial y de la cinética, respectivamente. Y el hamiltoniano es la energía total, es decir (para este ejemplo sencillo), la suma de las dos. Las derivadas parciales nos permiten obtener derivadas de cada variable por separado. Podemos así presentar las ecuaciones de Hamilton en su forma más general:

$$\frac{dp}{dt} = -\frac{\partial H}{\partial x}, \quad \frac{dx}{dt} = \frac{\partial H}{\partial p} \qquad (4.8)$$

Le prometo que no hay ninguna errata. De verdad hay un signo menos en la primera ecuación, pero no en la segunda. Hay razones para esto, pero se hallan profundamente ocultas en zonas de las matemáticas que no vamos a visitar aquí. (Puede buscar «topología simpléctica» si está interesado). Y sí, hay derivadas corrientes con letras d en la parte izquierda de cada ecuación, y derivadas parciales con símbolos ∂ en la parte derecha. Esto es porque posición y momento son funciones de tan solo una variable —el tiempo—, mientras que el hamiltoniano es una función de ambas cosas, tanto de la posición como del momento, de modo que necesitamos derivadas parciales que especifiquen qué estamos diferenciando con respecto a qué.

Es una forma elegante de abordar la cuestión. En la filosofía newtoniana, para cada parte del sistema había una ecuación de movimiento, que describía qué fuerzas actuaban en él. A posteriori, podíamos mostrar cómo se conservaban ciertas cantidades como «energía». En la filosofía hamiltoniana, en cambio, consignamos tan solo una ecuación, una expresión del hamiltoniano, la energía total como una función de posición y momento. A partir de ahí, podemos derivar todas las ecuaciones de movimiento necesarias. Esto sigue siendo verdad cuando tenemos sistemas

más complicados, con múltiples partes interactuando en todo número de formas. Seguirá habiendo uno solo hamiltoniano que describa todo el tinglado; cualquier cosa relacionada con la dinámica está implícita en esa expresión. La mecánica newtoniana, la mecánica lagrangiana y la mecánica hamiltoniana son equivalentes como teorías físicas, pero dependiendo de cuál de estas perspectivas adoptemos, puede hacernos la vida más fácil en según qué circunstancias.

Dije que no habría que llevarse deberes a casa, pero si le entraran ganas de hincar los codos, puede calcular la formulación explícita de las ecuaciones hamiltonianas para el oscilador armónico simple, en que la energía potencial es $V(x) = \frac{1}{2} \omega^2 x^2$. O incluso diseñar su propio hamiltoniano —¿qué tal dos osciladores, relacionados por algún tipo de interacción?—, y ver lo que sucede.

LOCALIDAD

¿A quién le importa? ¿Por qué estamos pasando por todo este esfuerzo matemático, solo para derivar un conjunto de ecuaciones que, a fin de cuentas, no hacen sino recapitular la mecánica newtoniana, pero con una formulación más intimidatoria?

Hay muchas razones para sentirse cómodo con la versión hamiltoniana de la mecánica, una de las cuales (no la menor) es que se revela indispensable al hacer la transición a la mecánica cuántica. Pero para nuestros objetivos presentes, pensar de acuerdo con este modelo nos ayuda a ver lo que el espacio tiene de especial.

Desde la perspectiva newtoniana, el «espacio» —es decir, el «espacio de posiciones», el conjunto de todas las posibles localizaciones— es claramente especial desde el comienzo. Podemos pensar en el «espacio de momentos», el conjunto de todos los posibles momentos, pero eso suena un poco abstracto, a diferencia de la naturaleza «tangible» del espacio de

posiciones. Vivimos en un espacio de posiciones, al fin y al cabo, no en un espacio de momentos.

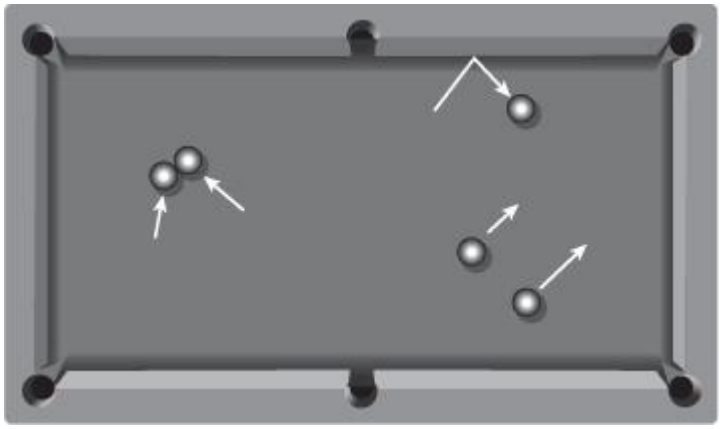
Pero cuando volvemos la vista hacia la perspectiva hamiltoniana, posiciones y momentos parecen estar en pie de igualdad, al menos de partida. Son las dos coordenadas del espacio de fases. El mismo hamiltoniano es una función tanto de x como de p. Y cuando observamos las ecuaciones de Hamilton, (4.8), posición y momento son casi simétricas (salvo por el signo menos). Por lo que a la estructura de la mecánica hamiltoniana concierne, el hamiltoniano podría ser cualquier función H(x, p). No hay nada en su formulación que distinga entre posición y momento para indicar: «esto es aquello en donde vivimos».

Así, en este escenario podemos plantear una pregunta que ni siquiera se les habría ocurrido a Newton o a sus inmediatos sucesores: ¿qué hay en el espacio que lo hace tan especial? ¿Por qué la posición y el momento nos parecen tan diferentes en la práctica, cuando se muestran tan iguales en las leyes físicas hamiltonianas? ¿Por qué vivimos en un espacio de posiciones y no en un espacio de momentos?

Lo que hace al espacio especial es que las interacciones son locales en cuanto a la posición. Los objetos interactúan directamente uno con otro cuando están en el mismo lugar, no cuando tienen el mismo momento (o cualquier otra cosa). Esto puede no ser obvio —¿acaso la fuerza gravitatoria del sol se expande a través del espacio hasta los planetas?—, de modo que tenemos que pensar un poco sobre ello.

Como físicos, es divertido hablar acerca del comportamiento de un sistema simple por sí mismo. Pero no tiene mucho sentido si no podemos también imaginar que observamos el sistema y comprendemos lo que hace. En última instancia, necesitamos describir múltiples sistemas interactuando, influyéndose y siendo influidos entre sí. Y una característica del universo en que vivimos es que los sistemas interactúan generalmente cuando cerca el uno del otro en el espacio. A esto es a lo que los físicos llamamos «localidad»: hacer algo en un punto particular del espacio solo afecta a otras cosas que se encuentran en ese punto o cerca de él. Finalmente tales efectos pueden transmitirse a otros lugares, pero eso lleva su tiempo.

Pensemos en unas bolas de billar rodando sobre una mesa. Por norma general se desplazan en línea recta (a no ser que apliquemos con el taco algún tipo de efecto) hasta que chocan con las bandas o con otra bola. Y esto sucede cuando coinciden en la misma localización espacial. No importa cuál sea su momento: dos bolas con momentos iguales (u opuestos, o con respecto a cualquier otra relación) no ejercerán una influencia especial la una en la otra.



Las bolas de billar interactuarán cuando entren en contacto con otra bola o con una banda en el mismo lugar del espacio. No sucede nada especial cuando dos bolas tienen el mismo momento.

Este hecho acerca de las interacciones en el mundo real se refleja en el hamiltoniano que termina describiendo sistemas del mundo real. En principio, el hamiltoniano de un sistema físico podría ser cualquier función de x y p. Pero en la práctica, encontramos sistemas descritos por hamiltonianos que adoptan la forma de (4.1): un término de energía cinética estándar, proporcional a p^2, y un término de energía potencial que depende únicamente de x, pero quizá con una forma funcional complicada.

Este patrón continúa cuando nos trasladamos a sistemas con múltiples partes móviles. Consideremos dos objetos definidos por las posiciones x1 y

x2 y los momentos p1 y p2. Casi todo el tiempo, el hamiltoniano terminará adoptando la forma:

$$H(x_1, p_1, x_2, p_2) = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + V(x_1, x_2 \qquad (4.9)$$

Reconocemos los dos primeros términos del lado derecho como las energías cinéticas individuales de los dos objetos. La energía potencial V(x1, x2) depende de algún modo de las posiciones de ambos objetos. Si hablamos de bolas de billar, el potencial será cero a menos que las bolas estén literalmente golpeando una con otra. Si hablamos de dos planetas que interactúan por medio de la gravedad, su energía potencial será cero cuando la distancia entre ambos sea muy grande, y será apreciable cuando se acerquen. Lo importante es que la distancia en el espacio es lo que determina la fuerza con que interactúan las cosas.

Estas son reglas generales que se aplican a sistemas típicos del mundo real, aunque hay algunas excepciones. La cuestión es que, aunque tengamos total libertad para imaginar todo tipo de hamiltonianos disparatados, los sistemas del mundo físico real no son tan arbitrarios. Los objetos tienen su propia energía cinética, e interactúan con otros objetos a través de la energía potencial, que depende de lo cerca que estén.

Esta es la razón por la que las posiciones son diferentes de los momentos, y por ende la razón también de que pensemos que vivimos en un espacio de posiciones y no en un espacio de momentos. En los hamiltonianos del mundo real, la posición es la variable en la que las interacciones son locales.

ACCIÓN A DISTANCIA

Hay aquí una sutileza que no deberíamos pasar por alto sin más. Con respecto a las bolas de billar, las interacciones son estrictamente locales en el espacio: las bolas no se influyen entre sí hasta el instante en que entran en contacto. Pero con por lo que respecta a la gravedad, la interacción sí que parece expandirse hasta cierta distancia, aunque vaya debilitándose a medida que aumenta la distancia. ¿Se trata aquí realmente de «localidad», tal como la hemos definido?

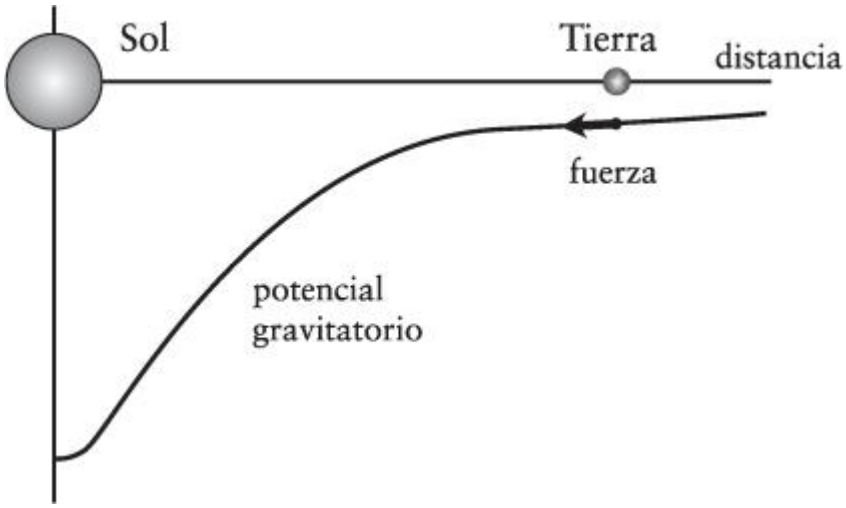
Buena pregunta. Newton y sus contemporáneos emplearon mucho tiempo reflexionando sobre la cuestión. Muchos estaban convencidos de que la fuerza de gravedad actuaba gracias a algún mediador, que tenía que haber algún tipo de sustancia entre los objetos, frente a la idea de que pudiera ejercerse libremente por el espacio. Habían aceptado implícitamente la idea de localidad, por lo que sospechaban que la «acción a distancia» no podía formar parte de una teoría correctamente formulada. El propio Newton estaba de acuerdo, calificando la noción de «tan grande y absurdo disparate, que no creo que haya ningún hombre con preocupaciones filosóficas mínimamente competente que pueda caer en tal error». Pero parecía funcionar, y si Newton era algo, era un hombre práctico. Así que comprender lo que en última instancia sucedía era algo que prefería «dejarlo a la consideración de los lectores».[2]

Le complacerá saber que las siguientes generaciones lo hemos considerado y en su mayor parte lo hemos comprendido. La cuestión es darse cuenta de que el espacio no está vacío, sino que está ocupado por campos, uno de los cuales es el responsable de la gravedad.

Un campo es básicamente una función del espacio mismo: para cada punto, el campo toma cierto valor. Dependiendo del tipo de campo del que estemos hablando, ese valor puede ser un número, un vector, o algo más complicado. Está el campo eléctrico, el campo magnético, el campo gravitatorio y varios otros. En física moderna, los campos son los componentes básicos fundamentales en la construcción de la realidad (hasta donde alcanza nuestro conocimiento actual).

Desde el punto de vista de los campos físicos, el sol no despliega mágicamente su alcance a través del espacio para ejercer su atracción gravitatoria sobre cada uno de los planetas. Más bien influye en el campo

gravitatorio, en el lugar que ocupa, de manera que el valor del campo en cada lugar influye en el valor de todos los lugares aledaños. Concatenando todos estos influjos —¡cálculo!— podemos determinar la fuerza de la atracción gravitatoria.



Campos y partículas son nociones complementarias. Una partícula se caracteriza por su localización en el espacio: está en ese único lugar, y en ningún otro. Un campo, en contrapartida, está por todas partes; todo campo tiene algún valor en cada una de las localizaciones espaciales. En cualquier punto del espacio que tomemos, cada campo experimenta la acción de otros campos, pero solo por los valores (y derivadas) de otros campos exactamente en el mismo punto del espacio. El valor de un campo en el punto A no tiene influencia directa sobre lo que sucede en el punto B; solo puede darse una influencia indirecta, mediada por cambios en el campo que pasa entre los dos puntos.

Fue nuestro viejo amigo Pierre-Simon Laplace quien determinó cómo podía funcionar esto en el caso de la gravedad. Planteó la idea de la existencia de un campo potencial gravitatorio ocupando el espacio. Este potencial se relaciona con la fuerza gravitatoria newtoniana de la misma forma en que la energía potencial sobre un paisaje de colinas se relaciona con la fuerza ejercida sobre una pelota en movimiento: la fuerza es la derivada del potencial.

Laplace proponía dos ecuaciones: una para determinar cómo responden los valores del campo a objetos macizos, y otra para mostrar cómo desplaza el campo a los objetos. En la imagen de Laplace, el sol crea una depresión en el potencial gravitatorio, que es cada vez menos pronunciada a medida que nos alejamos de toda esa masa. Este hecho se traduce en un declive en el campo, que los planetas experimentan como fuerza de gravedad. Las predicciones realizadas por la teoría de Laplace son completamente equivalentes a las de Newton, pero utilizando un vocabulario diferente para describirlas.

Las ecuaciones de Laplace son justamente locales: lo que sucede en un punto del espacio solo se ve afectado por lo que está sucediendo en las inmediaciones. Pero una característica que resulta desagradable para la sensibilidad moderna es que cualquier cambio acaecido en el campo gravitatorio se propaga instantáneamente a través del universo entero. Si usted coge una bola de un juego de bolos y la mueve de un lado para otro, el campo gravitatorio causado por la bola se modifica ligeramente; si Newton y Laplace hubieran estado en lo cierto, esa pequeña modificación podría detectarla al instante un detector lo suficientemente sensible situado en el otro extremo de la galaxia.

En nuestros días, después de haber sido instruidos por la teoría de la relatividad de Einstein, este tipo de transmisión instantánea de la información nos parece errónea. Las señales no pueden viajar más deprisa que la velocidad de la luz. De hecho, Einstein fue finalmente capaz de sustituir la concepción de Newton y Laplace acerca de la gravedad por su propia teoría, la relatividad general. La relatividad general insiste en postular un campo en el espacio-tiempo que explica la fuerza de la gravedad, pero matemáticamente es un poco más complicada. Por fortuna, es tan local como habríamos podido desear: el campo en sí mismo está influido únicamente por lo que sucede en la vecindad de cada punto, y cualquier modificación en el campo se transmite tan solo a la velocidad de la luz, o inferior.

Un último punto en torno a la localidad antes de seguir adelante. En física moderna, a la expresión «acción a distancia» suele añadírsele el adjetivo «fantasmagórica», parafraseando a Albert Einstein. Para cuando llegó el

siglo XX, el concepto de localidad estaba bien arraigado en la visión del mundo de los físicos, por lo que cualquier tipo de acción a distancia podría haber sido considerada fantasmagórica. La relatividad general consiguió liberar nuestras nociones acerca de la gravedad de tales preocupaciones. Pero la relatividad general no deja de ser una teoría clásica, en contraposición con la teoría cuántica. Lo que veía Einstein era que la llegada de la mecánica cuántica abría la puerta a un nuevo tipo de acción a distancia, resultante de las mediciones de las partículas cuánticas entrelazadas. El efecto que él propuso es real y ha sido verificado experimentalmente. Pero la naturaleza es lo bastante sutil como para que ninguna información útil pueda jamás viajar a una velocidad superior a la de la luz. Así que la mecánica cuántica nos deja en una situación extraña, pues tenemos que utilizar conceptos no locales para explicar lo que observamos, pero no podemos beneficiarnos de ellos para enviar señales a través del espacio.

1 A las unidades se las llama a veces dimensiones, de ahí el nombre de «análisis dimensional». Pero reservaremos el nombre «dimensiones» para el número de direcciones independientes en que pueden moverse los objetos.

2 Carta de Isaac Newton a Richard Bentley del 25 de febrero de 1692. The Newton Project, < newtonproject.ox.ac.uk/view/texts/normalized/THEM00258/ >.

CINCO

TIEMPO

Justificábamos nuestra consideración acerca del «espacio» destacando que era el escenario en el que acontecen las cosas. Pero «acontecer» implica una transformación en el tiempo: hay algo que no ha sucedido, que luego está sucediendo y que finalmente ha sucedido. Si no hubiera tiempo, no habría movimiento, ni evolución, ni cambio. Cualquier cosa de interés en el universo debe su existencia al paso del tiempo.

A veces hablamos como si el tiempo fuera algo misterioso e inefable. De acuerdo con la famosa formulación de San Agustín: «Si nadie me lo pregunta, sé lo que es el tiempo. Si quiero explicarlo a quien me lo pregunta, ya no lo sé». Es decir, no hay ningún misterio en el concepto de tiempo tal como lo utilizamos en la práctica ordinaria. Si alguien nos dice: «Nos vemos a las ocho», o «este pódcast dura noventa minutos», sabemos exactamente qué quieren decirnos. El universo —el espacio y todas las cosas incluidas en él— acontece, una y otra vez, con ligeros cambios, a lo largo de una secuencia de momentos. El tiempo es una etiqueta puesta en esos momentos y una forma de medir la duración de algún conjunto de momentos.

Pero por mucho que el tiempo sea menos misterioso de lo que a veces suponemos, ciertamente tiene algunos aspectos misteriosos. ¿En qué son similares el tiempo y el espacio, y qué los diferencia? ¿Por qué el pasado es diferente del futuro? Todas estas son excelentes preguntas, a las cuales podemos aportar respuestas en diferente grado de certeza.

TIEMPO Y ESPACIO

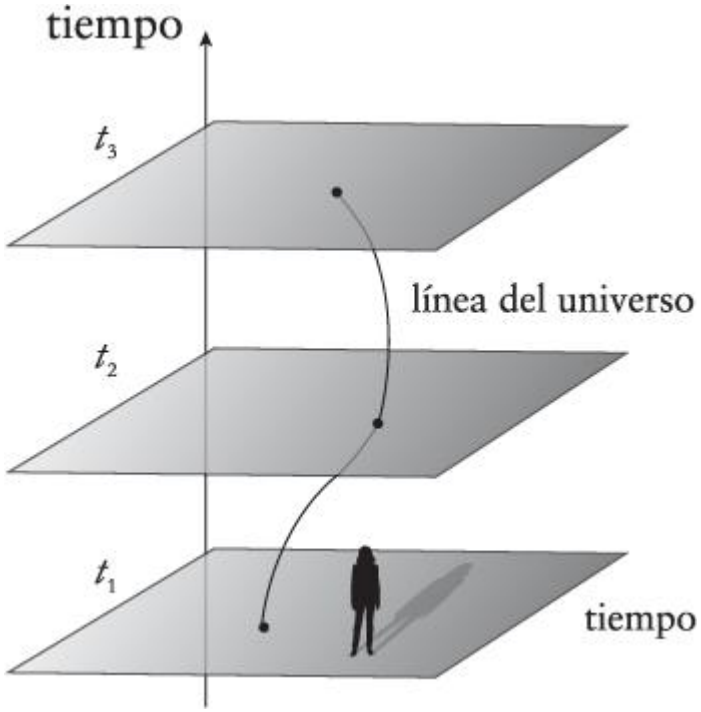
El tiempo es difícil de captar porque en ciertos sentidos es como el espacio, pero en otros no lo es. Detengámonos a considerar, en primer lugar, el sentido en que el tiempo es como el espacio.

El tiempo forma parte del modo en que nos ubicamos a nosotros mismos en el universo. Si quiere quedar con alguien para tomar un café, tiene que especificar tanto el dónde como el cuándo. Podemos codificar esta información mediante cuatro números: tres para determinar una localización en el espacio tridimensional y uno para concretar un momento singular en el tiempo. En la práctica, solemos dar por implícita parte de esta información —diremos: «en la cafetería del centro», dando por sentado que nuestro amigo conoce tanto la dirección como la altura sobre el nivel del suelo—; pero en principio está ahí. Y si le dijéramos a alguien dónde encontrarnos, pero no cuándo, o viceversa, no serviría de mucho.

Una vez que tenemos asumida la consideración del tiempo y del espacio como formas de ubicarnos a nosotros mismos en el universo, es natural que los agrupemos en un mismo paquete al que llamamos «espacio-tiempo». Es posible que haya oído decir que el concepto de espacio-tiempo fue acuñado a comienzos del siglo XX, en el marco de la teoría de la relatividad. Como dato histórico, puede ser cierto. Pero incluso en la física newtoniana, antes de la llegada de la relatividad, uno podía reunir el espacio y el tiempo, formando un conglomerado espacio-temporal. La diferencia estriba únicamente en que la división del espacio-tiempo en «espacio» y «tiempo» está enteramente fijada en la física newtoniana —hay una única forma correcta de hacerlo—, mientras que en la teoría de la relatividad, los diferentes observadores dividirán por lo general el espacio-tiempo en espacio y tiempo de forma diferente. En la física newtoniana, tiempo y espacio son, cada uno de ellos, absolutos por derecho propio, mientras que en la relatividad, como sabe, son relativos. (En particular, son relativos con arreglo a nuestra elección de un marco referencial en el espacio-tiempo).

Aunque la física newtoniana no nos obligue a reunir el espacio de tiempo en un conglomerado, nos invita amablemente a hacerlo. Una vez hecho, tiene sentido trazar una gráfica, una representación simplificada del universo que incluya tanto las dimensiones espaciales, como la dimensión única del

tiempo. Tales diagramas de espacio-tiempo constituyen una parte fundamental del juego de herramientas conceptuales del físico moderno. Por regla general, suelen representarse con el espacio expandido horizontalmente y el tiempo orientado en sentido vertical, y con el pasado hacia la parte de abajo del diagrama y el tiempo aumentando a medida que nos desplazamos hacia arriba.[1]



En un diagrama de espacio-tiempo, un objeto no estará representado por un simple punto que denote su posición, sino por una línea de universo que irá del pasado al futuro, y que definirá sus diversas posiciones en diferentes momentos en el tiempo. Nuestro cuerpo ocupa un volumen tridimensional en el espacio en cualquier momento del tiempo, pero nuestra historia vital dibuja un gusano de cuatro dimensiones que se extiende por el espacio-tiempo. Todo científico, en un momento u otro, ha visualizado su propia línea de universo desde el momento de su nacimiento hasta el momento presente.

LONGITUD Y DURACIÓN

La otra similitud obvia entre el tiempo y el espacio es que podemos medirlos. Los caminos que recorren el espacio tienen una longitud mensurable, como los intervalos de tiempo tienen una duración mensurable.

En la práctica habitual, podemos medir cosas cuando poseemos un patrón fiable con el que compararlas. En el caso de la longitud, tenemos reglas. Tomemos dos barras sólidas de la misma longitud, que protegeremos de cualquier deterioro, y dejemos pasar un tiempo. Quizás hayamos guardado una de las barras en un armario, quizá la otra haya viajado a algún otro lugar, para acabar regresando. Cuando la segunda barra vuelve, volvemos a compararlas y comprobamos que siguen teniendo la misma longitud. Es la fiabilidad y reproductibilidad de esta comparación lo que nos proporciona un patrón útil de medida. Podemos coger una de las barras, llamarla «metro patrón» y medir todas las demás longitudes con respecto a él. En última instancia, el origen de la existencia de tan maravilloso artefacto puede buscarse en las leyes subyacentes de la física. Las barras están hechas de átomos, y los átomos poseen medidas fijas, básicamente, establecidas por constantes físicas tales como las masas y cargas eléctricas de los electrones y los núcleos atómicos. No deje que nadie sustraiga (por roce accidental o por malicia) ningún átomo de su metro patrón, y ya lo tiene.

Con la duración viene a suceder lo mismo, solo que la medimos con relojes en lugar de barras. Un reloj puede concebirse como algo que cambia con el tiempo según una pauta que es fiable y reproducible en comparación con otros relojes. Es posible que esto parezca una definición circular, ya que en ella estamos utilizando la noción de reloj para definir el reloj, pero no estamos haciendo nada diferente de lo que hacíamos al medir barras: son útiles en tanto conservan la misma longitud unas con respecto de las otras.

La cuestión es que el universo real está lleno de relojes: sistemas que realizan algún tipo de movimiento de un modo regular y predecible los unos con respecto de los otros. Esto es una característica de las leyes de la física que nos permite medir el paso del tiempo. (Es divertido trata de imaginar cómo serían las cosas si no hubiera objetos-relojes en el universo).

Un ejemplo clásico lo constituye la rotación de la Tierra y su órbita alrededor del sol. Podemos tener una razonable confianza en que, en el transcurso de un año, la Tierra habrá rotado alrededor de su eje un poco más de 366 veces.[2] Esta fiabilidad es la que hace a los buenos relojes.

Otro ejemplo es el del oscilador armónico simple. Si un oscilador es verdaderamente armónico —la fuerza de restitución es exactamente proporcional a su desviación del estado de equilibrio— tardará la misma cantidad precisa de tiempo en experimentar una oscilación completa, sea cual sea la amplitud de la oscilación. El oscilador se mueve más rápido cuando oscila mucho, y más lento cuando lo hace solo un poco, de forma que el periodo es el mismo en cualquier caso. Esto hace del principio de los osciladores armónicos un excelente punto de partida para la construcción de relojes.

Los osciladores armónicos de precisión son difíciles de encontrar en la naturaleza. Suele haber siempre algún pequeño efecto que hace que el sistema físico real no sea estrictamente armónico. Pero por fortuna hemos comprobado, ya que los sistemas oscilatorios tienden a ser aproximadamente armónicos cuando su oscilación es lo bastante pequeña. Es famosa la historia que se cuenta de Galileo, recogida por su primer biógrafo, según la cual advirtió que una lámpara que colgaba del techo de la catedral de Pisa oscilaba con un periodo fijo de tiempo. Ello le dio la idea de que un péndulo podía servir de principio para construir un reloj preciso. Años más tarde (cuando ya se había quedado ciego), Galileo recuperó esta idea y, con la ayuda de su hijo, propuso cómo podría funcionar un reloj así. No llegaron a completar el proyecto, pero poco después Christiaan Huygens desarrolló el primer diseño funcional.

Un péndulo, de todos modos, no es un oscilador armónico perfecto: cuando sus oscilaciones son demasiado grandes, no tienen el mismo periodo que cuando son pequeñas. Para garantizar su exactitud, necesitaremos fundamentar nuestro reloj en un péndulo que oscile tan solo un poco, para aprovechar el hecho de que todos los osciladores se comportan armónicamente con amplitudes pequeñas. Robert Hooke inventó finalmente un artilugio llamado escape de áncora, que permitía la construcción de

relojes de péndulo funcionales que oscilaban apenas unos grados en cada oscilación.

EVOLUCIÓN EN EL TIEMPO

Hasta aquí la parte fácil, las similitudes del tiempo en relación con el espacio. Pero el tiempo no es exactamente como el espacio. Intuitivamente, el tiempo parece acontecer de un modo que no es completamente equiparable con el del espacio. Sentimos que el tiempo fluye, o como si nosotros pasáramos por él. No se nos ocurre describir el espacio como algo que fluye a nuestro alrededor, o que nosotros fluimos por él. Para nuestra intuición indocta, al menos, el espacio parece una colección de cosas (de puntos, uno por cada localización), mientras que el tiempo se asemeja a una especie de proceso.

Como tantas concepciones intuitivas, este sentimiento de que el tiempo y el espacio son cosas por completo diferentes no sale de ninguna parte. En el tiempo, por ejemplo, las condiciones presentes en un momento evolucionan sin discontinuidad hacia las condiciones del momento siguiente. Pensemos en la segunda ley de Newton, que equipara la fuerza al producto de la masa por la aceleración (que a su vez es la segunda derivada de posición):

$$\vec{F} = m\vec{a} = m\frac{d^2\vec{x}}{dt^2} \qquad (5.1)$$

O también, consideremos las ecuaciones de Hamilton para el tiempo derivadas de posición y momento:

$$\frac{dp}{dt} = -\frac{\partial H}{\partial x}, \quad \frac{dx}{dt} = \frac{\partial H}{\partial p} \qquad (5.2)$$

En cualquiera de los dos casos, estas relaciones valdrán para cada uno de los momentos del tiempo. Determinan lo que sucede en otras ocasiones.

Nada parecido funciona igual con respecto al espacio. Si se encuentra con una pesada roca —demasiado pesada para moverla con cualquier medio disponible a su alcance—, hay grandes posibilidades de que la roca siga ahí al cabo de un pequeño lapso de tiempo. Pero si se encuentra con una pesada roca en una ubicación determinada, no hay ninguna ley que diga que tendrá que haber otra roca (o más de la misma roca) en una ubicación próxima en el espacio. Al considerar posiciones diferentes en el espacio, las condiciones pueden cambiar drásticamente de un punto a otro. Pero al considerar momentos en el tiempo, una condición fluye directamente a la siguiente.

LA FLECHA DEL TIEMPO

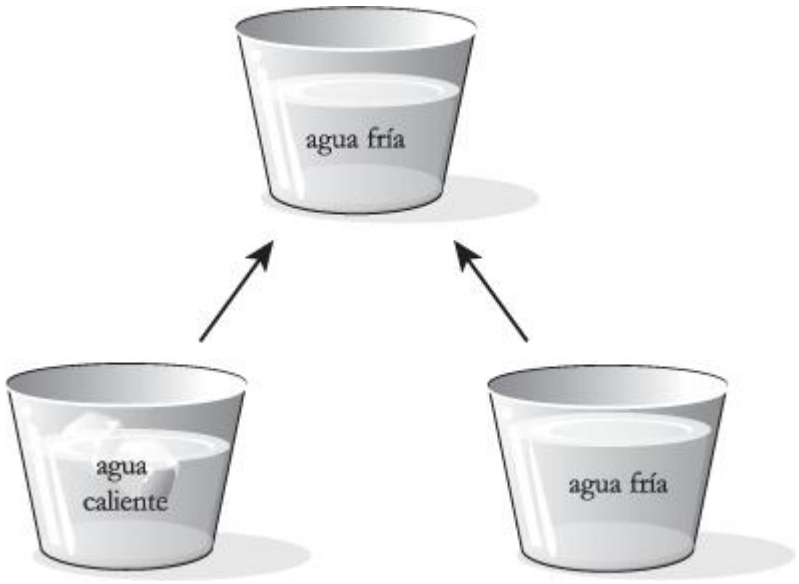
Intuitivamente también, ese «fluir» del tiempo parece tener una pronunciada direccionalidad. El tiempo parece moverse del pasado al futuro, y no al revés. Esta característica constituye la famosa flecha del tiempo, llamada así por el astrofísico británico Arthur Eddington en 1927.

La flecha del tiempo es fundamentalmente asimétrica: desde el punto de vista del «ahora», pasado y futuro tienen propiedades por completo diferentes. El pasado es inamovible, está en los libros, mientras que el futuro está como abierto, a nuestra disposición. Tenemos recuerdos y testimonios del pasado: fotografías, fósiles, libros de historia, artefactos del presente que nos proporcionan cierto conocimiento definido acerca de lo que sucedió en algún periodo de tiempo anterior. No hay nada de esto con respecto al futuro. Podemos intentar predecirlo, pero no tenemos ninguna imagen de él. Nos gusta pensar que nuestras elecciones aquí y ahora pueden influir en cómo será el futuro, pero la mayoría de nosotros admitirá que nuestras elecciones de hoy no influyen en mañana.

La flecha forma parte de un modo tan indeleble de nuestra noción del tiempo y de su funcionamiento, que los primeros pensadores ni siquiera habrían pensado en ello como en algo que requiriera explicación. Por supuesto que el pasado y el futuro son diferentes, ¿cómo, si no? Son cosas fundamentalmente diferentes.

El advenimiento de la mecánica clásica hizo que la flecha del tiempo adquiriera unos tintes un tanto misteriosos. En este marco aprendimos que la información se conserva. A partir del estado presente de un sistema aislado, podemos tanto predecir su futuro, como retrotraernos a su pasado. Esta característica se conoce como reversibilidad: si las leyes de la física implican que un sistema en estado A en un tiempo 1 evoluciona a un estado B en un tiempo 2, por esas mismas leyes se infiere que si el sistema estuviera en el estado B en el tiempo 2, debe haber estado antes en el estado A en el tiempo 1. Tanto podemos adelantar el reloj, como atrasarlo con igual adecuación.

La reversibilidad no es algo que resulte manifiesto en nuestra experiencia cotidiana. Si a un vaso de agua caliente le echamos hielo, al cabo de un tiempo habrá pasado a ser un vaso de agua a temperatura ambiente. Pero un vaso cuya agua esté de inicio a temperatura ambiente también será un vaso de agua a temperatura ambiente al cabo de un tiempo (es decir, no habrá cambiado en ese tiempo). Por tanto, si nos encontramos con un vaso de agua a temperatura ambiente, no podemos saber si antes era un vaso de agua caliente con hielo en su interior, o si ha sido un vaso de agua a temperatura ambiente durante todo el tiempo.



Así pues, la mecánica clásica encierra una reversibilidad secreta: las leyes implícitas de la física son reversibles aunque nuestra experiencia del mundo que nos rodea ciertamente no lo sea. Decimos a veces que la dinámica «microscópica» o «fundamental» es reversible, pero que la dinámica «macroscópica» o «manifiesta» no lo es. Aquí, «microscópica» significa «constituida por un pequeño número de partes móviles», más que «de tamaño pequeño». La mecánica newtoniana reversible describe muy bien el comportamiento de los planetas en el sistema solar, aunque sean bastante grandes. La cuestión es entonces: ¿por qué el mundo macroscópico parece irreversible?

INVERSIÓN DEL TIEMPO

Antes de profundizar en el origen de la irreversibilidad, deberíamos examinar el concepto de simetría de inversión temporal, que, aunque suene muy similar, es diferente en el fondo. Puede resultar un poco confuso, pero vale la pena aclararlo, ya que oirá a algunas personas hablar de él con la pretensión de que está en relación con la flecha del tiempo (y no lo está, en realidad).

Una simetría es una transformación que cambia de sitio algunas de las variables del sistema, pero deja lo esencial intacto. La transformación de inversión temporal T implica tomar la variable temporal t y multiplicarla por −1 para invertir su dirección:

$$t \rightarrow -t \qquad (5.3)$$

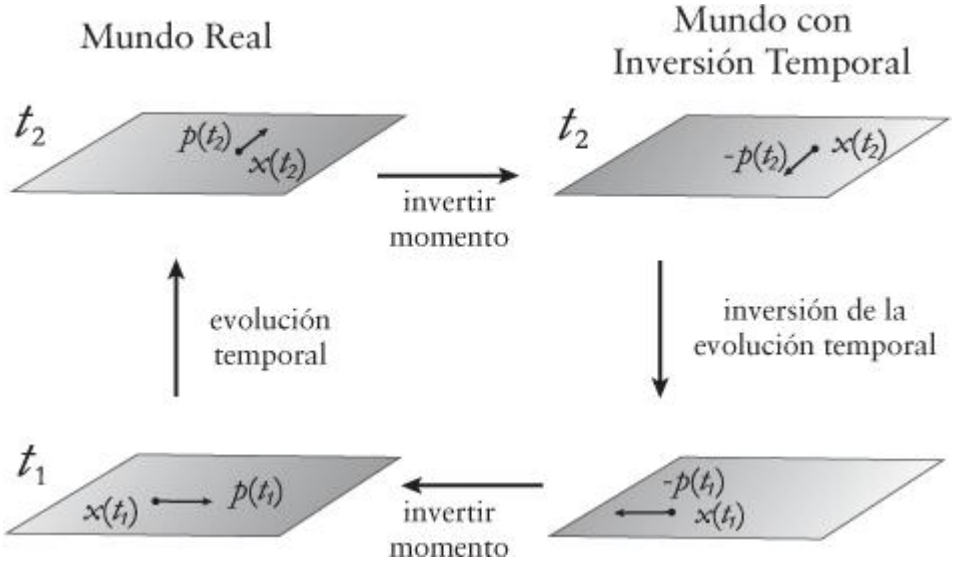
En esta transformación, las leyes de la física clásica son invariantes, permanecen sin cambio. Esta característica de la invariancia de la inversión temporal está estrechamente relacionada con la reversibilidad, pero no es del todo lo mismo.

Pensemos en la segunda ley de Newton, tal como estaba expresada en (5.1). La masa del objeto no cambia con la inversión del tiempo. Ni tampoco la fuerza, al menos en ejemplos conocidos como la fuerza de gravedad o la pendiente de una colina. ¿Y qué sucede con la aceleración, ? La posición no cambia. El intervalo infinitesimal de tiempo dt sí lo hace; dado que t → −t, también tenemos dt → dt. (Si invertimos la dirección del tiempo, también invertimos la dirección en que el tiempo se incrementa). Pero la aceleración implica el cuadrado de esto: = d / dt = d² / dt², por lo que tenemos: (dt)² → (dt)² = (dt)². Así, la aceleración permanece inalterada por la transformación (5.3). De ahí que la integridad de la segunda ley de Newton sea invariante con la inversión del tiempo.

Lo mismo es aplicable a la versión hamiltoniana de la mecánica clásica, pero de un modo ligeramente más divertido. Hay dos «trucos» que deberá tener en mente. En primer lugar, cuando realice la inversión temporal, deberá enviar el momento a menos él mismo: . Puede que piense que es una obviedad, puesto que el momento es el producto del tiempo por la velocidad, y las velocidades son inversas cuando llevamos el tiempo hacia atrás. Pero el momento no se define como el producto de la masa por la velocidad en la perspectiva hamiltoniana: es una variable independiente que resulta ser igual a la masa por la velocidad en trayectorias que resuelven las ecuaciones de movimiento. En cualquier caso, tiene que invertir su signo si

quiere obedecer a la simetría de inversión temporal. (Esto es ya un indicio de que hay algo sutil en todo este asunto).

La segunda peculiaridad que hay que retener en la mente es que el hamiltoniano es la energía cinética más la energía potencial, y la energía cinética suele expresarse como . Dado que el momento se eleva al cuadrado, la energía cinética misma permanece invariante cuando invertimos el signo del momento.



En otras palabras, las ecuaciones fundamentales de la mecánica clásica no conocen la diferencia entre moverse hacia el pasado y moverse hacia el futuro.

REVERSIBILIDAD Y CPT

«Reversibilidad» e «invariancia de la inversión temporal» son dos ideas que suenan igual, pero entre las que hay una sutil diferencia. La reversibilidad dice que la información se conserva con el transcurso del tiempo, de forma que los sistemas no olvidan cuál era el estado del que partieron; la

invariancia de la inversión temporal dice que las leyes de la física ofrecen la misma visión tanto si van hacia delante como hacia atrás. La reversibilidad se revela así más fundamental: siempre que la dinámica sea reversible, podremos diseñar alguna noción de simetría de inversión temporal.

Podemos abordar la cuestión dando un rodeo. Si frecuenta a las personas equivocadas —en este caso, a los físicos de partículas—, es posible que le digan que en el mundo de las partículas elementales, la inversión del tiempo no es una simetría de física fundamental, a fin de cuentas. Incluso hemos llevado a cabo experimentos que demuestran la vulneración del principio de invariancia de la inversión temporal. Bajo las circunstancias adecuadas y precisas, puede usted tomar un conjunto de partículas en cierto estado A, hacerlas evolucionar en el tiempo hasta un diferente estado B, invertir luego todos los momentos para hacerlos «retroceder en el tiempo», y que no terminen precisamente en el estado A, como podíamos haber esperado.

No es esta la única simetría que podía esperar encontrar en la naturaleza, pero que en realidad resulta fallida. Además de la inversión temporal, tenemos también la inversión espacial, conocida habitualmente como paridad y simbolizada con la letra P. Una transformación de paridad revierte las tres dimensiones del espacio a la vez; es equivalente a mirar un sistema en un espejo, de tal forma que la escritura aparece al revés y los tornillos de rosca derecha se ven como tornillos de rosca izquierda. Las partículas elementales suelen tener espín, y una transformación de paridad haría que la dirección de su espín fuera la inversa. En principio se pensó que la paridad era un buen ejemplo de simetría en la naturaleza, pero en una serie de experimentos realizados en la década de 1950, C. S. Wu y otros mostraron que se vulneraba.

Hay también una simetría llamada conjugación de carga (C). Muchas especies de partículas elementales llevan asociados diferentes tipos de antipartículas. La antipartícula del electrón recibe su propio nombre, el positrón, pero en la mayoría de los casos recurrimos al prefijo anti-, de modo que la antipartícula de un neutrino es un antineutrino. La conjugación de carga intercambia partículas por sus correspondientes antipartículas. Al igual que sucede con T y con P, también C resulta vulnerada por ciertas interacciones entre partículas. La conjugación de carga puede parecer una

especie taxonómica diferente de la inversión espacial o temporal, pero la dirección del tiempo está íntimamente conectada con la distinción entre materia y antimateria; existe un determinado sentido en que las antipartículas son matemáticamente equivalentes a partículas que viajaran hacia atrás en el tiempo.

Colofón: individualmente, C, P y T resultan todas ellas vulneradas en la física de partículas elementales. Pero la combinación de transformaciones, CPT, se conserva. Hay buenas razones teóricas para considerar esto verdad, además de haberse confirmado experimentalmente. Si tomamos una disposición de partículas en estado A y las hacemos evolucionar en el tiempo hasta un estado B, podemos construir la imagen especular de B, intercambiar partículas con antipartículas, invertir todos los momentos, retroceder en el tiempo y volver a obtener la imagen especular. El resultado será nuestro estado original A. Tal es la simetría CPT en acción.

Todo ello sugiere la siguiente atrevida estrategia: ¿podríamos, tal vez, definir simplemente la «inversión del tiempo» como lo que acabamos de llamar CPT? En otras palabras, ¿podríamos definir una nueva operación, llamémosla T', realizando en primer lugar una inversión temporal naíf (T), luego una transformación de paridad (P) y finalmente una conjunción de carga (C), y considerar esta nueva operación T' = CPT «aquello que la inversión temporal significa realmente»? Esta nueva versión manipulada de la inversión temporal constituiría una simetría natural, a fin de cuentas.

Ciertamente podemos hacerlo. De hecho, siempre podemos definir un operador de inversión temporal cambiando la dirección del tiempo más las alteraciones apropiadas de otras variables, de modo que el resultado sea una buena simetría… si la teoría implícita es reversible. La reversibilidad implica cierta noción de simetría de inversión temporal, aunque pueda no parecer precisamente la simetría naíf que supusimos en un principio.

Recuerde que incluso en la mecánica hamiltoniana convencional tenemos que invertir la dirección del momento para llevar a cabo la inversión temporal, pero no ofrecimos realmente una justificación basada en principios para dar este paso. Ahora sabemos por qué: definir la inversión temporal para que sea una buena simetría de la dinámica implicará generalmente algún tipo de manipulación, además de invertir $t \rightarrow t$. En

ocasiones estos pasos adicionales pueden parecer naturales e incluso inevitables, como invertir , de tal forma que nos limitamos a añadirlo a la definición de «inversión temporal» y nos olvidamos de ello. Otras veces pueden parecer ad hoc, como introducir la paridad y la conjunción de carga, lo cual nos deja mascullando algo así como «se ha vulnerado la inversión temporal, pero sigue conservándose una simetría estrechamente relacionada». Lo que cuenta, desde el punto de vista de la física, e independientemente de las definiciones por las que optemos, es que siempre existe algún tipo de simetría con tal de que la dinámica sea reversible.

De modo que la invariancia de inversión temporal naíf T no tiene nada que ver con la flecha del tiempo. Esta característica de la física de partículas, por importante que sea por sí misma, deja la reversibilidad intacta. La noción de flecha del tiempo nace del hecho de que el mundo macroscópico no se muestra como reversible, aunque el mundo microscópico sí lo sea, aparentemente.

ENTROPÍA

¿Qué es, entonces, aquello que es responsable de la irreversibilidad, y por ende de la flecha del tiempo? La respuesta última reside en el hecho de que la entropía de un sistema cerrado, incluido el universo entendido en su totalidad, tiende a aumentar con el tiempo. Muchas veces se define la entropía en términos muy generales como el desorden o desorganización de un sistema. Un mazo de cartas ordenadas perfectamente tiene una entropía baja, mientras que tras barajarlo al azar tiene una entropía elevada. Esto puede valer para la mayor parte de nuestros objetivos, pero podemos ser un poco más precisos.

Para el demonio de Laplace, aquel que rastreaba los estados del mundo en sus detalles más microscópicos, todo parecería ser reversible. Recordemos además que el demonio conoce también enteramente el pasado y el futuro. El demonio no hace distinciones entre «recordar el pasado» y «predecir el futuro». Para el demonio, no existe la flecha del tiempo en la realidad.

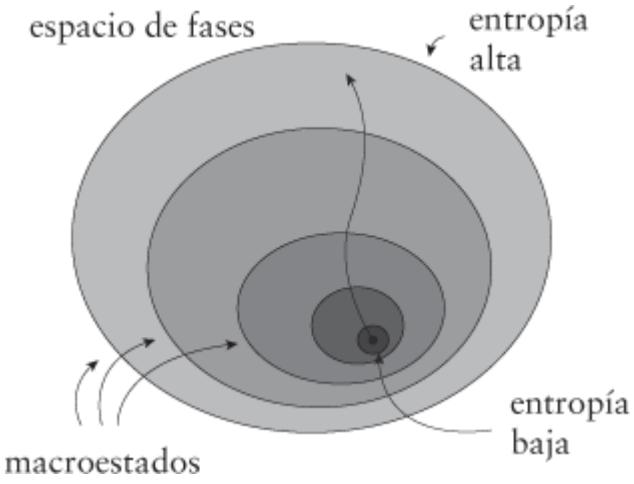
Pero ninguno de nosotros es el demonio de Laplace. Ni siquiera por aproximación. Somos criaturas humanas finitas, con unas capacidades de observación y de cálculo irremediablemente limitadas. Si tenemos problemas para recordar un número de teléfono, qué decir de aprender las posiciones y momentos de 6×10^{23} o más partículas.[3] No lo sabemos todo acerca del exacto estado del mundo, y ni siquiera vemos el exacto estado del mundo cuando lo observamos. Cuando miramos en torno a una habitación, vemos mesas, sillas, personas; no medimos con precisión la posición y el momento de cada una de las partículas elementales de que están hechas.

Lo que hacemos, en vez de eso, es lo que llamamos granulado grueso: agrupamos un puñado de estados específicos de un sistema en una descripción simple, utilizamos esa descripción para entender lo que está pasando y realizamos las mejores predicciones que podemos. Cuando describimos un contenedor con gas o una taza de café, podemos especificar la temperatura, la presión y la velocidad del fluido en cada punto del contenedor —el macroestado del sistema—, pero siempre habrá muchas otras disposiciones de átomos y moléculas — microestados— que encajen con la descripción. Sin embargo, no necesitamos conocer el microestado exacto para saber que el gas se expandirá hasta llenar el contenedor, o que el café caliente se enfriará hasta alcanzar la temperatura ambiente al cabo de un tiempo. Podemos hacer buenas predicciones basándonos simplemente en el conocimiento de los macroestados que tenemos a nuestra disposición.

Dar una definición precisa de macroestado es una tarea delicada, pero la idea básica sería: «el conjunto de todos los microestados que ofrecen el mismo aspecto para un observador macroscópico». Algunos macroestados, como el que se da cuando un conjunto de moléculas de gas se distribuye uniformemente en el interior de un contenedor, se corresponden con un gran número de posibles macroestados. Otros macroestados, en cambio, como cuando acontece sin más que las moléculas de gas se agrupan todas en un rincón del contenedor, se corresponden con un número relativamente pequeño de posibles microestados.

Fue el físico austríaco Ludwig Boltzmann quien, durante la década de los años 1870, tuvo la brillante idea de utilizar esta estructura para sugerir un

modo de entender la entropía. La idea de entropía había sido introducida ya antes, en el siglo XIX. La contribución de Boltzmann consistió en conectar propiedad intrínsecamente macroscópica con sus sustentáculos microscópicos. En concreto, propuso que la entropía estaba relacionada con el número de microestados en cada macroestado.[4] Desde esta perspectiva, tiene sentido que la entropía tienda a aumentar con paso del tiempo. Un macroestado de baja entropía se corresponde nada más que con un pequeño número de posibles microestados, mientras que un macroestado de alta entropía se corresponde con un gran número de posibles microestados. Si partimos de un estado de baja entropía y nuestro sistema se aleja siguiendo cierta dirección genérica en el espacio de fases, deberíamos esperar que la entropía aumente, simplemente porque hay más maneras (habitualmente muchas, muchas más maneras) de presentar una entropía alta que una entropía baja.



Esta es la segunda ley de la termodinámica, según la cual, en un sistema cerrado la entropía o bien aumentará o permanecerá constante, nunca disminuirá espontáneamente. (La primera ley se refiere básicamente a la conservación de la energía). Muchos sistemas no son cerrados, en la medida en que pueden interactuar con el mundo exterior. En ese caso, la entropía puede ciertamente disminuir; cuando ponemos una botella de champaña en el frigorífico, se enfría y su temperatura baja. Pero notará que el frigorífico expulsa calor por la parte de atrás. Podemos seguir diciendo que la entropía, considerada en su totalidad, aumenta. Este es el motivo por el que no puede

refrigerar una habitación enchufando en ella un frigorífico y dejando la puerta de este abierta. Las leyes de la termodinámica no se lo permitirán.

La aportación de Boltzmann ofrece una conexión entre el microcosmos de las partículas y el demonio de Laplace, y entre el microcosmos del granulado grueso, la aproximación, y la información incompleta. El hecho de poder decir algo útil acerca de los sistemas físicos, mientras permanecemos ignorantes de tantos detalles sobre sus sistemas microscópicos, es un ejemplo de una característica conocida como emergencia. Dejaremos por ahora de lado la importante cuestión de por qué el mundo admite descripciones emergentes de nivel superior, pero es algo que debemos retener en la mente.

LA HIPÓTESIS DEL PASADO

La entropía tiende a aumentar porque hay más modos (microestados) de tener alta entropía que de tener baja entropía. Quizás haya advertido que esta explicación introduce subrepticiamente una presunción oculta: que todo comenzó con un nivel bajo de entropía. Esta idea, conocida simplemente como hipótesis del pasado, colabora con la conexión micro/macro de Boltzmann para poner en marcha la segunda ley de la termodinámica.

Podemos arrojar un poco de luz recurriendo una vez más a la comparación entre tiempo y espacio. Mientras que claramente vemos que hay una flecha del tiempo, no existe nada parecido a una «flecha del espacio». Podemos distinguir una diferencia entre «arriba» y «abajo», pero eso es únicamente porque vivimos en el entorno de un objeto de fuerte influencia —la Tierra — y experimentamos la fuerza de atracción de su gravedad. Si fuera usted un astronauta flotando con un traje espacial, no percibiría ninguna diferencia apreciable entre arriba y abajo, izquierda y derecha, adelante y atrás.

Pero, un momento. El hecho es que estamos en la Tierra, y aquí existe una flecha que distingue el arriba del abajo. Lo que sucede es que tal hecho no forma parte de las leyes de la física, sino que se trata de un hecho accidental propio de nuestro entorno local. Es debido a la particular configuración de la materia en nuestras inmediaciones, que toma la forma de un gran planeta que ejerce su influjo a su alrededor. ¿Podría suceder esto mismo con el tiempo?

Podría, y de hecho así es. Dado que las leyes subyacentes de la física son reversibles, si lo único que supiéramos fuera que la entropía presente del universo es baja, podríamos predecir que será más alta en el futuro, pero también que fue más alta en el pasado. Estaríamos viviendo en cierto momento histórico especial en que se da el hecho anómalo de que la entropía es baja.

Nadie cree que esto sea verdad. La simetría fundamental entre pasado y futuro la rompe una condición límite de la historia de nuestro universo, o al menos de la parte observable del mismo. Dicha condición límite establece que la entropía inicial era verdaderamente muy baja.

Nuestro universo está en expansión, y hace unos catorce mil millones de años todo lo que ahora vemos estaba comprimido en un estado muy caliente y muy denso, pronto a expandirse. Si nos remontamos atrás todo ese tiempo, la mejor teoría que poseemos en la actualidad (la relatividad general de Einstein) postula una «singularidad» de infinita densidad, etiquetada con el nombre de «Big Bang» («Gran Explosión»). El modo adecuado de entender todo esto no es pensar que realmente se dio tal singularidad, sino que la propia teoría se derrumba al llegar a ese punto. Algún día poseeremos una mejor comprensión de la gravedad y del universo en expansión, que esperamos nos revele lo que sucedió en el momento al que llamamos Big Bang.

Por ahora, podemos contentarnos con describir lo que sucedió justo después de ese instante, pues acerca de ello tanto la teoría como la observación nos han dicho muchas cosas. El universo consistía en un denso y caliente plasma, distribuido uniformemente por el espacio.

Esto puede sonar a una configuración de alta entropía: si aquí en la Tierra tenemos un contenedor de gas en un laboratorio, la configuración de más alta entropía posible es aquella en que el gas está distribuido uniformemente. Pero hay una diferencia fundamental entre esta situación y cuando estamos hablando del universo considerado en su totalidad, que es tan grande y masivo, que la gravedad es determinante. La gravedad actúa perturbando dicha uniformidad, atrayendo a la materia hacia las regiones más densas y vaciando las zonas menos densas. Esto es precisamente lo sucedido durante los siguientes pocos miles de millones de años, a medida que las estrellas, las galaxias, los cúmulos estelares han ido congregándose y apartándose de la uniformidad inicial. Durante todo este proceso, la entropía ha ido aumentando. Cuando la gravedad es importante, hay más posibilidades de que las partículas se agrupen de forma desigual que uniforme. De modo que el estado inicial del universo tenía una entropía enormemente más pequeña de la que habríamos podido esperar si hubiera sido (de algún modo) elegido al azar de entre todas sus posibles configuraciones.

No sabemos por qué el universo partió de una configuración con baja entropía. Después de todo, se supone que las configuraciones de baja entropía constituyen un caso especial, pues se corresponden con microestados cuyo número es mucho menor que el de los estados con entropía alta. Esta es una cuestión para la cosmología, que dejaremos de lado por ahora.

Lo que nos importa es que nuestro universo observable partió de un estado de baja entropía y que la entropía ha ido en aumento desde entonces. Tal es el origen último de la dirección del tiempo. Así como hay una flecha del espacio en nuestras inmediaciones porque vivimos en el entorno de la Tierra, existe también una flecha del tiempo porque vivimos cerca (relativamente hablando) del Big Bang.

Para ser sinceros, queda mucho por explicar. La flecha del tiempo se manifiesta de muchas formas: por la memoria, por la causalidad, por el envejecimiento, y por muchas más. Hemos estado hablando de la flecha del tiempo «termodinámica», con la idea de que subyace en última instancia en todos los demás aspectos que implican direccionalidad, pero no lo hemos

demostrado con rigor. Tales cuestiones se encuentran todavía en un estado abierto a la investigación.

PRESENTISMO, ETERNALISMO, POSIBILISMO

Hemos representado una imagen según la cual la flecha del tiempo no estaría inscrita dentro de las leyes fundamentales de la física, sino que se trataría más bien de un epifenómeno, un subproducto del hecho de que el universo partiera de una condición de baja entropía, y de que los estados de baja entropía tiendan a evolucionar a estados de entropía en constante aumento. La direccionalidad o flujo del tiempo no forma parte de la arquitectura fundamental de la realidad, que no hace distinción entre evolucionar hacia el pasado o el futuro, ni entre diferentes instantes del tiempo.

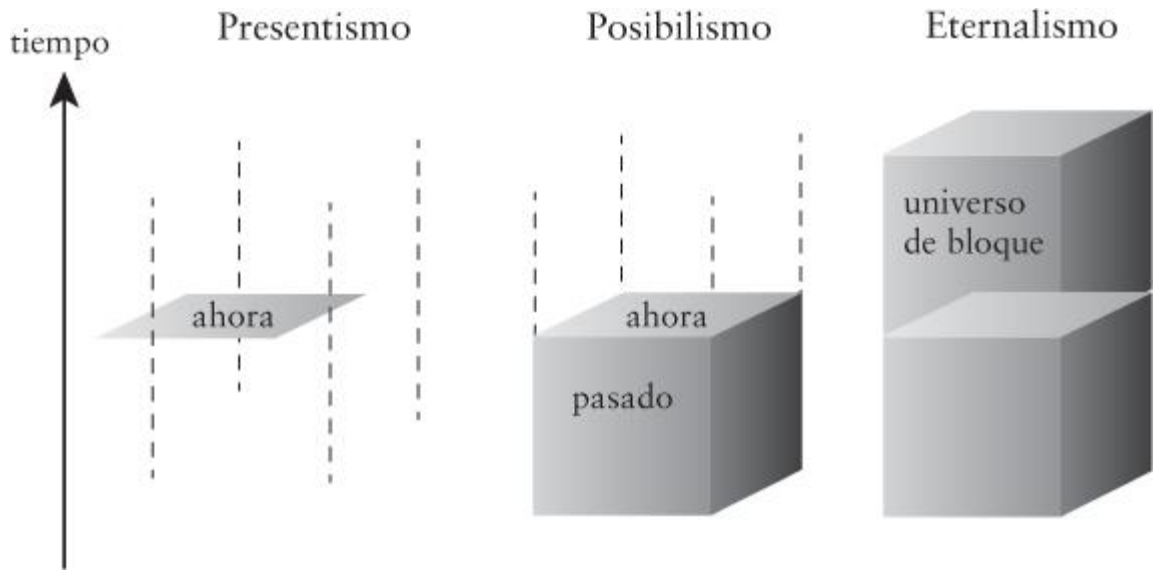
Sería este un buen momento para hacer notar que no todo el mundo está de acuerdo. La historia que acabamos de contar es probablemente lo más cercano que tenemos a una visión «estándar» entre los científicos y filósofos que han pensado sobre estas cuestiones, pero no está universalmente aceptada. La física está bastante bien establecida, sobre este punto, pero posiciones filosóficas alternativas dignas de consideración (y que podrían redundar algún día en una mejor comprensión de la física, si resultaran estar en lo cierto).

Es preciso preguntarnos, una vez más, qué entendemos por «real». En nuestra discusión en torno al espacio, distinguíamos entre la idea de que el espacio es una entidad separada por derecho propio, o si no es más que una forma de organizar las relaciones que se establecen entre los objetos. Pero no se nos ocurría pensar que algunas localizaciones en el espacio fueran reales y otras no. Cuando la cuestión es el tiempo, este pensamiento no solo se nos ocurre, sino que es primario y central. Tradicionalmente, ha sido más común la concepción de que tan solo el momento presente del tiempo, el «ahora», es verdaderamente real. El pasado ya se fue, de acuerdo con este

punto de vista, mientras que el futuro todavía no ha llegado. Por tanto, ni pasado ni futuro son reales en el mismo sentido en que lo es el presente.

La concepción según la cual únicamente es real el momento presente, se conoce como presentismo; suena lógico. Puede confrontarse con el eternalismo, noción según la cual todos los momentos del tiempo son reales por igual, cosa que implícitamente hemos asumido en algunas de las discusiones anteriores. También se conoce al eternalismo como teoría del universo de bloque, ya que considera que el mundo real es el bloque entero de cuatro dimensiones del espacio-tiempo.

Presentismo y eternalismo no son las únicas opciones sobre el tapete. Quienes se sienten atraídos por el presentismo, es posible que se muestren también sensibles a la diferencia existente entre el modo en que suponemos el pasado y el futuro. No podemos cambiar el pasado, de eso no hay duda, mientras que sí que pensamos que nuestras acciones aquí y ahora pueden tener influencia en el futuro, aún por definir. De ahí que una tercera y tentadora posibilidad sea el posibilismo, o concepción del «presente en crecimiento». En este caso, consideramos reales tanto el pasado como el presente, mientras que negamos tal condición al futuro.



No vamos a adjudicarle la razón definitiva a ninguna de estas perspectivas en liza, pero sí podemos reconocer que existen sensatos motivos detrás de cada una de ellas. El presentismo está más cerca de nuestra visión intuitiva, precientífica, del mundo. Los eternalistas consideran las leyes de la física, advierten que son reversibles y que conservan la información, y que además no distinguen ningún momento privilegiado como «ahora»; de todo ello, piensan, se desprende que podemos conceder igual estatus al pasado, al presente y al futuro. Podemos explicar nuestras impresiones de que el tiempo fluye y de que el «ahora» posee un algo más real que otros momentos como consecuencias de la flecha del tiempo, la cual es consecuencia a su vez del aumento constante de la entropía.

A todo ello los presentistas negarán con la cabeza, lamentando que sus amigos eternalistas no salgan de un embrollo para meterse en otro a la hora de explicar lo que tienen delante de las narices. ¿No sería más fácil, dicen, explicar el fluir del tiempo postulando que el tiempo fluye realmente? ¿Y que solo el «ahora» es real, y el tiempo es nuestro modo de hablar del proceso de cambio, y no meramente una coordenada de no se sabe qué realidad cuadridimensional absolutamente intangible? Ciertamente, dicen los posibilistas, y desde el momento en que nos tomamos nuestras intuiciones en serio, no seamos tan miopes como para tratar el pasado y el futuro en pie de igualdad.

SER Y DEVENIR

Todas estas son versiones modernas de discusiones antiguas. Muchas veces se ha hecho remontar el rastro del presentismo hasta el filósofo griego Heráclito, quien acentuaba la primacía del cambio: no podemos sumergirnos dos veces en el mismo río, ya que en cada momento es un río diferente. Una perspectiva más eternalista puede encontrarse, una generación más tarde, en Parménides, quien concibió un universo eterno sin fin. Parménides, en otras palabras, priorizaba el ser, mientras que Heráclito situó el concepto de devenir en la cima de la metafísica.

La distinción entre ser y devenir se refleja no solo en la dicotomía eternalismo/presentismo, sino en el modo mismo en que concebimos las propias leyes de la naturaleza. Según cierta escuela de pensamiento, las leyes no son más que formas convenientes de resumir lo que acontece en el mundo. Podríamos imaginarnos describiendo el universo mediante una lista de lo que sucede en cada instante del tiempo, pero es más sencillo (aunque no por ello deja de ser incalculablemente difícil) describir lo que sucede en un único instante, para decir a continuación: «y el resto de instantes se asocian con este por las leyes de la naturaleza». Esta perspectiva se ha llamado humeanismo, por el filósofo escocés David Hume, aunque no es obvio que Hume hubiera podido calificarse de humeano en este sentido.

La alternativa se conoce como antihumeanismo, presumiblemente porque no ha habido ningún filósofo que se distinguiera en la defensa de esta perspectiva con la suficiente convicción como para que su nombre quedara asociado a ella, por mucho que cuente con multitud de partidarios.[5] Los antihumeanos piensan que las leyes de la naturaleza no se limitan a describir el mundo, sino que actúan a cada instante para hacer que el mundo exista. Las leyes, en otras palabras, no son meran acompañantes: tienen una existencia independiente y son responsables de que el mundo sea como es.

El humeanismo encaja muy bien con la mentalidad eternalista, mientras que los antihumeanos es más probable que sean presentistas. Para apreciar de qué modo se relacionan estos puntos de vista, consideremos las propuestas contrafactuales: afirmaciones acerca de cómo habría sido el mundo de haber sido un poco diferente. («Me habría saltado el aperitivo si hubiera sabido que la carta de postres era tan irresistible»). Puede considerarse que los contrafactuales se postulan acerca de otros mundos posibles, como versiones sensatas de la realidad que, simplemente, sucede que no son la realidad en que nos encontramos. Recurrimos a ellas constantemente: si no pensáramos en la posibilidad de otros mundos factibles, no podríamos hablar en términos de elección, responsabilidad, moral, u otras ideas a las que por lo general estamos bastante apegados.

Pero un momento. Para los humeanos, todo cuanto existe es el universo real. Lo que llamamos «leyes de la naturaleza» son simplemente compendios de patrones útiles que apreciamos dentro de este universo.

Entonces, ¿cómo podría un humeano hablar siquiera de contrafactuales? O como mínimo, ¿cómo podría imaginar coherentemente un universo diferente, pero con las mismas leyes de la física? Los humeanos piensan que las leyes se limitan a describir el mundo, no que lo gobiernen. Los antihumeanos no tienen esta preocupación. Para ellos, las leyes no son un mero compendio de lo que sucede, sino que gobiernan lo que sucede. A partir de ahí, es fácil extender mentalmente este poder a otros mundos posibles, con otras condiciones iniciales pero con las mismas leyes.

No obstante, los antihumeanos se enfrentan al reto de explicar qué significa decir que las leyes de la física «gobiernan» lo que sucede en cada instante. Si supiéramos todo lo que ha acontecido en el universo, sería difícil señalar con exactitud qué dato adicional nos faltaría. Los humeanos tienen siempre la ligera sospecha de que los antihumeanos siguen anclados en una anticuada perspectiva cripto-aristotélica, en que las partes del mundo tienen esencias y naturalezas. Quizá sea lo correcto pensar que las cosas suceden sin más, y que nuestra tarea consiste en describir qué son esas cosas.

1 Al menos entre las personas especializadas en relatividad y espacio-tiempo. Los científicos de la computación suelen representar el incremento del tiempo hacia abajo, mientras que los físicos de partículas lo hacen en una orientación de izquierda a derecha. (O podría a ser de derecha a izquierda, si hablan una lengua que se escriba en esa dirección).

2 ¿Por qué no 365? Porque aunque la Tierra no rotara, el sol saldría y se pondría una vez al año, debido a que la Tierra no habría dejado de cumplir una órbita a su alrededor. Y ese movimiento sería en sentido contrario al que vemos a causa de la rotación. Terminamos viendo una puesta de sol menos al año, con respecto al número de rotaciones completas de la Tierra.

3 El moderno ordenador portátil en que escribo esto tiene 64 GB de RAM, capacidad de memoria suficiente para almacenar con gran aproximación posiciones y momentos de unas 109 partículas. Necesitaríamos aproximadamente mil millones de ordenadores como este para almacenar la información en un sistema macroscópico relativamente pequeño, suponiendo que empleara toda su memoria para la tarea. Es algo que no va a suceder en un futuro próximo.

4 Más específicamente, la entropía de un macroestado es proporcional al logaritmo del número de microestados incluidos en él. Véase Apéndice A para una discusión en torno a los logaritmos.

5 Según una encuesta reciente, el 31% de filósofos adoptan una perspectiva humeana acerca de las leyes de la naturaleza, un 54% son antihumeanos y un 15% respondieron: «otra». D. Bourget y D. Chalmers (2021),

«Philosophers on Philosophy: The 2020 PhilPapers Survey», < philarchive.org/archive/BOUPOP-3/ >.

SEIS

ESPACIO-TIEMPO

En la física newtoniana, se habría sentido alentado a considerar juntos el espacio y el tiempo, como una entidad individual de cuatro dimensiones, el espacio-tiempo, en la cual se localizan los acontecimientos. Pero nada le habría obligado a hacerlo. El espacio y el tiempo tienen sus propias identidades independientes, que nadie ha confundido nunca. Fue con la teoría de la relatividad, formulada a comienzos del siglo XX, cuando se hizo inevitable hablar del espacio-tiempo. En la relatividad, ya no es válido decir que el espacio y el tiempo poseen significados objetivos y separados. Lo que realmente existe es el espacio-tiempo, y escindirlo en espacio y en tiempo es una mera y útil convención humana.

Una de las razones principales por las que la relatividad tiene fama de ser de difícil comprensión es que todas nuestras intuiciones nos habitúan a pensar en el espacio y el tiempo como cosas separadas. La experiencia que tenemos de los objetos es que estos poseen una extensión en el «espacio», lo cual se nos antoja un hecho bastante objetivo. En última instancia, esta concepción nos basta, porque por lo general viajamos a través del espacio a velocidades ampliamente inferiores a la de la luz, de modo que la física prerrelativista funciona bastante bien.

Pero este desajuste entre intuición y teoría nos lleva de hecho a dar el salto a una perspectiva espacio-temporal en cierto modo intimidatoria. Para empeorar las cosas, las exposiciones de la teoría de la relatividad con frecuencia abordan la cuestión de abajo arriba, partiendo de nuestras concepciones cotidianas del espacio y del tiempo, y mostrando cómo resultan alteradas en el nuevo contexto de la relatividad. Nos cuentan historias acerca de barras de metal que se estiran y de relojes

distorsionados, todo ello técnicamente pertinente, pero que puede oscurecer la belleza y simplicidad de los conceptos subyacentes.

Vamos a hacer las cosas de un modo un poco diferente. Nuestra ruta hacia la relatividad especial será más bien descendente, tomándonos en serio la idea de un espacio-tiempo unificado desde el arranque, y viendo lo que ello implica, en lugar de partir de nociones más familiares en apariencia, para luego apreciar cómo se ven modificadas en este el nuevo contexto. Vamos a tener que tensar un poco el cerebro, pero como resultado obtendremos una comprensión mucho más profunda de la perspectiva relativista acerca de nuestro universo.

EL NACIMIENTO DEL ESPACIO-TIEMPO

Suele atribuirse el desarrollo de la teoría de la relatividad a Albert Einstein, pero lo que él hizo fue coronar un edificio teórico que había estado construyéndose desde que James Clerk Maxwell unificara la electricidad y el magnetismo en una sola teoría acerca del electromagnetismo en la década de los años 1860. La teoría de Maxwell explicaba lo que es la luz —una onda oscilante en el seno de los campos electromagnéticos—, y parecía conferir un lugar de privilegio a la velocidad a la que viaja la luz. La idea de un campo existente por sí mismo no era totalmente intuitiva para los científicos de la época, por lo que era natural investigar la cuestión de qué era lo que «ondeaba» realmente en una onda de luz.

Varios físicos investigaron la posibilidad de que la luz se propagara al través de un medio al que apodaron «éter luminífero». Pero nadie pudo encontrar pruebas que demostraran la existencia de tal éter, por lo que se vieron obligados a inventar razones cada vez más complicadas de por qué esta sustancia era indetectable. La contribución de Einstein de 1905 consistió en señalar que la existencia de dicho éter ser había vuelto completamente innecesaria, y que podíamos entender mejor las leyes de la física sin él. Todo lo que teníamos que hacer era aceptar una concepción

enteramente nueva acerca del espacio y del tiempo. (De acuerdo, era mucho aceptar, pero resultó valer de verdad la pena).

La teoría de Einstein terminó conociéndose como teoría de la relatividad especial, o simplemente relatividad especial. Su artículo fundamental llevaba por título: «Sobre la electrodinámica de los cuerpos en movimiento», y en él se propugnaban nuevas formas de pensar acerca de la longitud y la duración. Explicaba el papel especial que desempeña la velocidad de la luz, postulando que en el universo existe una velocidad límite absoluta —la velocidad a la que viaja la luz a través del espacio vacío — y que todo el mundo descubriría al calcularla que es siempre la misma, sin importar el movimiento del propio observador. Para que todo ello fuera concebible, tuvo que modificar en consecuencia nuestras nociones acerca del tiempo y del espacio.

Pero no llegó a tanto como a preconizar la reunión del espacio y el tiempo en un simple espacio-tiempo unificado. Este paso lo dio su antiguo profesor universitario, Hermann Minkowski, en 1907. El marco de la relatividad especial se conoce hoy como espacio-tiempo de Minkowski.

Una vez planteada la noción de espacio-tiempo como un continuo unificado de cuatro dimensiones, podemos comenzar a hacernos preguntas acerca de su forma. El espacio-tiempo, ¿es plano o curvo, estático o dinámico, finito o infinito? El espacio-tiempo de Minkowski es plano, estático e infinito. Einstein trabajó durante una década para comprender cómo podía integrarse la fuerza de gravedad en su teoría. Su gran avance final fue discernir que el espacio-tiempo podía ser dinámico y curvo, y que los efectos de su curvatura responden a lo que usted y yo experimentamos como «gravedad». Los frutos de esta inspiración son lo que ahora llamamos relatividad general.

Así pues, la relatividad especial es la teoría de un espacio-tiempo fijo, plano, sin gravedad; la relatividad general es la teoría de un espacio-tiempo dinámico, en el que la curvatura da lugar a la gravedad. Ambas se cuentan entre las teorías «clásicas» aunque sustituyan algunos de los principios de la mecánica newtoniana. Para los físicos, «clásico» no significa «no relativista», sino «no cuántico». Todos los principios de la física clásica, incluida la posibilidad de pensar en las cosas en términos de dinámica

hamiltoniana o del principio de mínima acción se mantienen perfectamente intactos en el contexto relativista, aunque hayamos llegado a una noción algo más sofisticada del espacio y el tiempo.

DOS NOCIONES DE TIEMPO

A menudo se presenta la relatividad especial en términos de «contracción de la longitud» y «dilatación del tiempo». Tales conceptos son perfectamente respetables, pero reflejan una tendencia anticuada a tratar el espacio y el tiempo separadamente, como fundamentales ambos, pero más expandidos de lo que pensamos de ordinario.

La verdad es más profunda. Tenemos que estar dispuestos a desprendernos de nuestro apego prerrelativista por la separación entre espacio y tiempo, y permitir que estos se fundan en el marco unificado espacio-temporal. La mejor manera de lograrlo es pensar, con mayor detenimiento aún, acerca de lo que queremos decir cuando hablamos de «tiempo». Y el mejor modo de hacer esto es recordar, una vez más, lo que pensamos acerca del espacio.

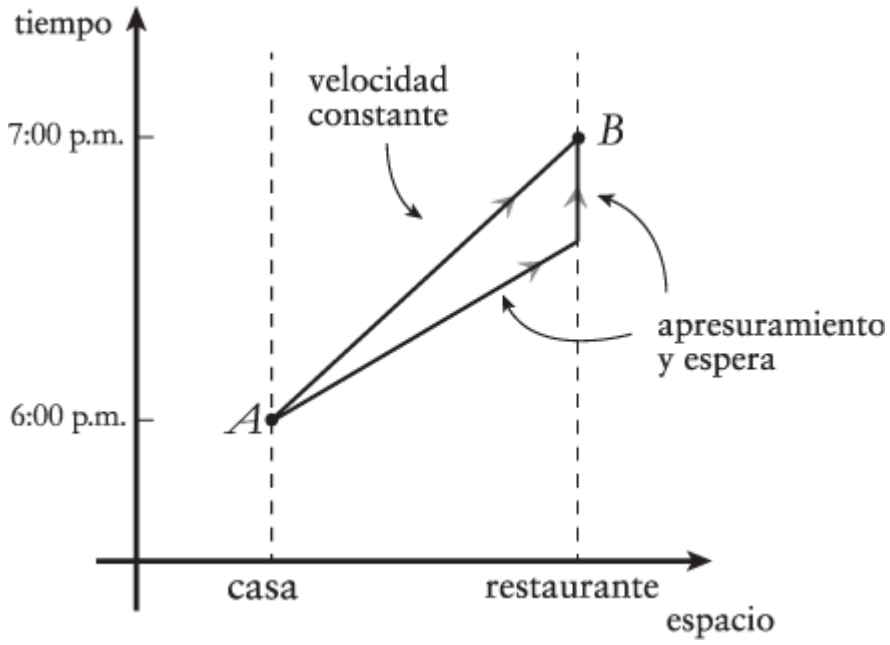
Considere dos lugares en el espacio como, por ejemplo, su casa y su restaurante favorito. ¿Cuál es la distancia entre ambos?

Bueno, eso depende, pensará inmediatamente. Porque está la distancia «a vuelo de pájaro», es decir, la que imaginamos si pudiéramos trazar un trayecto perfectamente rectilíneo entre ambos puntos. Pero está también la distancia que recorrería en un desplazamiento en el mundo real, con la limitación de tener que tomar vías públicas, cruzar aceras, evitar edificios y otros obstáculos que pudieran surgir a lo largo del camino. La distancia que cubra siempre será más larga que la distancia ideal «a vuelo de pájaro», puesto que la línea recta es la distancia más corta entre dos puntos.

Considere ahora dos sucesos en el espacio-tiempo. En la jerga técnica de la teoría de la relatividad, un «suceso» es simplemente un punto en el universo, definido por su localización tanto en el espacio como en el

tiempo. Un suceso, llamémoslo A, puede ser «en casa a las seis de la tarde», y otro suceso B puede ser «en el restaurante a las siete de la tarde». Con estos sucesos en mente, piense en un desplazamiento entre A y B. Por mucho que corra, no estará en B antes: si llega al restaurante a las 18:45 h, tendrá que esperar hasta las 19:00 h para alcanzar el suceso en el espacio-tiempo que habíamos designado como B.

Podemos ahora preguntarnos, tal y como habíamos hecho en el caso de la distancia espacial entre nuestra casa y el restaurante: ¿cuánto tiempo ha transcurrido entre estos dos sucesos?



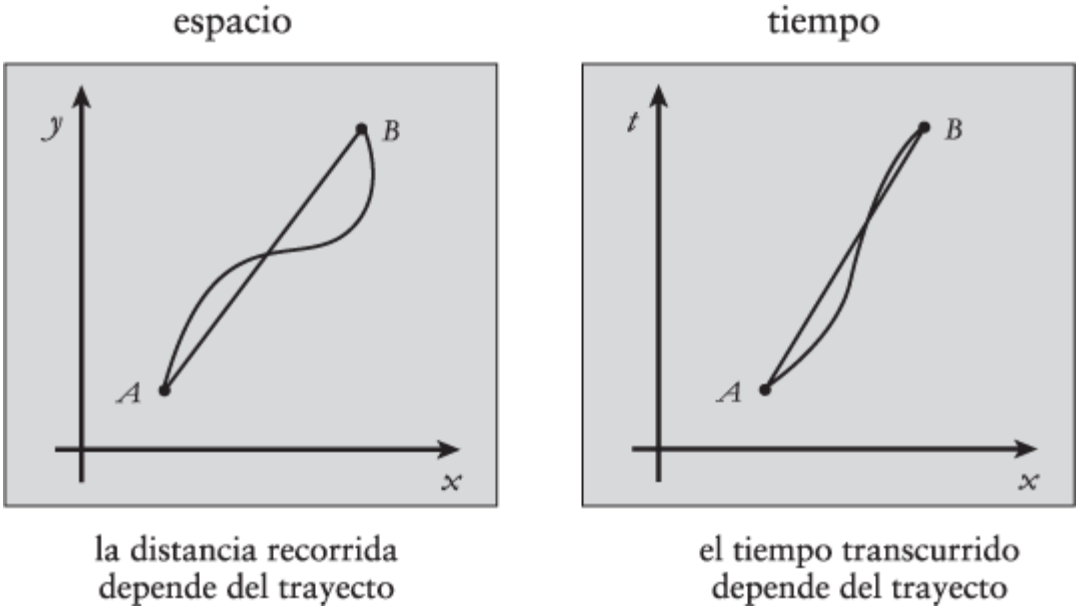
Quizá piense que se trata de una pregunta trampa. Si un suceso es a las 18:00 h y el otro a las 19:00 h, transcurre una hora entre ambos, ¿no es así?

No es tan sencillo, diría Einstein. En una anticuada concepción newtoniana del mundo, sí, por descontado. El tiempo es absoluto y universal, y si el tiempo entre dos sucesos es de una hora, no hay nada más que decir.

La relatividad cuenta una historia diferente. Tenemos ahora dos nociones distintas acerca de lo que queremos decir cuando decimos «tiempo». La primera sería la noción de tiempo como coordenada espacio-temporal. El

espacio-tiempo es un continuo de cuatro dimensiones y, si queremos especificar posiciones en él, es conveniente asignar un número al que llamamos «tiempo» a cada punto contenido en él. En términos generales, esto es lo que tenemos en mente cuando pensamos en «18:00 h» o «19:00 h». Estos números se refieren a valores señalizados en unas coordenadas espacio-temporales; son etiquetas que nos ayudan a ubicar sucesos. Se supone que todo el mundo entiende lo que queremos decir cuando decimos «nos encontramos en el restaurante a las 19:00 h».

Pero, dice la relatividad, del mismo modo que la distancia a vuelo de pájaro es por lo general diferente de la distancia que realmente recorremos entre dos puntos en el espacio, la duración del tiempo que experimentamos a lo largo de nuestra línea de universo no será en general la misma que la de la coordenada universal de tiempo. Experimentamos una cantidad de tiempo que podría medirse mediante un reloj que lleváramos en nuestro interior durante el desplazamiento. Este es el tiempo propio del recorrido. Y la duración medida por el reloj, al igual que la distancia medida por el cuentakilómetros del coche, dependerá del trayecto que sigamos.



la distancia recorrida depende del trayecto

el tiempo transcurrido depende del trayecto

Vemos aquí uno de los aspectos de lo que significa decir «el tiempo es relativo». Podemos pensar tanto en un tiempo común en términos de una

coordenada espacio-temporal, como en un tiempo personal que experimentamos individualmente a lo largo de nuestro camino. Y el tiempo es como el espacio: estas dos nociones no tienen por qué coincidir. (Como señaló el historiador Peter Galison, no es ninguna coincidencia que Einstein trabajara en una oficina de patentes en Suiza, en una época en que los rápidos viajes por ferrocarril obligaban a los europeos a pensar en qué hora era en otras ciudades del continente, por lo que perfeccionar los relojes se convirtió en un importante reto tecnológico).

TRAYECTORIAS INERCIALES REPRESENTAN TIEMPOS MÁS LARGOS

Tiene que haber algún sentido en que el tiempo no sea como el espacio, de lo contrario simplemente hablaríamos de un espacio de cuatro dimensiones, en lugar de singularizar el tiempo etiquetándolo de forma diferenciada. Y no nos referimos por ahora a la flecha del tiempo: de momento hemos regresado a un mundo simple con pocas partes en movimiento, en el que la entropía y la irreversibilidad no son cosas que nos preocupen.

La diferencia es la siguiente: en el espacio, una línea recta define la distancia más corta entre dos puntos; en el espacio-tiempo, en cambio, un camino recto ofrece el tiempo más largo transcurrido entre dos sucesos. Este giro de la distancia más corta al tiempo más largo es lo que distingue al tiempo del espacio.

Por «camino recto» en el espacio-tiempo queremos significar tanto una línea recta en el tiempo como una velocidad constante de desplazamiento. En otras palabras, una trayectoria inercial, sin aceleración. Determinemos dos sucesos en el espacio-tiempo: dos ubicaciones en el espacio con sus correspondientes instantes en el tiempo. Un viajero podría realizar el recorrido entre ambos en línea recta a velocidad constante (a cualquiera que fuera la velocidad requerida para llegar en el instante prefijado), o bien podría desplazarse desviándose aquí y allá, siguiendo un camino no inercial a lo largo del trayecto. La ruta irregular siempre implicará más distancia

espacial, pero menos tiempo propio transcurrido que la opción del camino rectilíneo.

¿Por qué? Porque así lo dice la física. O, si lo prefiere, porque el universo está hecho así. Quizás algún día descubramos una razón más profunda de por qué las cosas tenían que ser de este modo, pero en nuestro estado actual del conocimiento, es uno de los presupuestos fundamentales sobre los que se asienta la física, y no una conclusión que hagamos derivar de principios más profundos. Las líneas rectas en el espacio constituyen la distancia más corta posible; los caminos rectos en el espacio-tiempo conllevan el tiempo más largo posible.

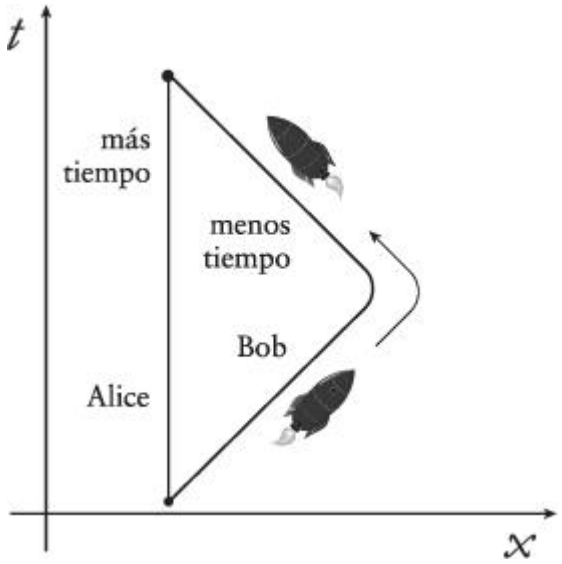
Puede parecer contrario a la intuición que los caminos de mayor distancia requieran de menor tiempo propio. Estamos de acuerdo. Si fuera intuitiva, no habríamos necesitado a Einstein para dar con la idea.

EL EXPERIMENTO MENTAL DE LOS GEMELOS

Hay un ejemplo muy ilustrativo de este principio, conocido habitualmente como la paradoja de los gemelos, si bien, indiscutiblemente, de ningún modo es una paradoja. Se trata de una característica no intuitiva de la física con la que no solemos encontrarnos en nuestra vida diaria, porque por supuesto solemos movernos a velocidades muy inferiores a la de la luz. No es muy acertado atribuir palabras como «paradoja» a fenómenos que tienen perfecto sentido, pero que nos parecen inusuales cuando se nos presentan por primera vez.

Consideremos dos gemelos, Alice y Bob. En realidad, ni siquiera tienen por qué ser gemelos, dos personas que se desplacen por trayectorias diferentes a través del espacio-tiempo experimentarán diferentes cantidades de tiempo, sean o no parientes, pero la situación es más gráfica si son de la misma edad. Mientras Alice permanece en la Tierra, Bob monta en una nave espacial extremadamente avanzada, sale disparado a una velocidad próxima a la de la luz, da media vuelta y vuelve raudamente junto a Alice.

Imaginaremos que la aceleración de la nave de Bob es prácticamente instantánea y que él resulta indemne a pesar de las enormes fuerzas gravitatorias implicadas en el proceso. Ignoraremos también detalles como el hecho de que la Tierra se desplace (lentamente) por el espacio, o que Alice esté emplazada en un campo gravitatorio. Los experimentos mentales nos permiten acogernos a este tipo de aproximaciones al estilo de la vaca esférica.



En consonancia con esta situación, Alice se ha movido siguiendo una línea recta a través del espacio-tiempo, sin llegar a moverse siquiera en el espacio. La línea de universo de Bob lo lleva del mismo suceso inicial al mismo suceso final, pero siguiendo un camino claramente no rectilíneo, expresado por esa curva en el punto medio. (Su trayectoria después de partir habría sido rectilínea si no hubiera dado media vuelta, pero en tal caso no habría estado de regreso junto a Alice en ese mismo suceso espacio-temporal posterior, y no habría habido forma de comparar sus edades).

Lo que descubren los gemelos al volver a reunirse es que Bob ha experimentado menos tiempo propio que Alice. No es tan viejo como ella, y si llevaba un reloj consigo, señalará menos tiempo transcurrido. Para hacer balance, si Bob viaja a una velocidad constante en relación con la Tierra de un 99 % la velocidad de la luz, por cada año que él experimente, Alice

experimentará en torno a siete años. (Podrá hacer usted mismo este cálculo más adelante). Si Bob tardó unos pocos años en completar su viaje, al regresar habrá descubierto que su hermana gemela es considerablemente más vieja que él.

No hemos enviado (todavía) a ningún astronauta en un viaje de ida y vuelta a una velocidad cercana a la de la luz para poner a prueba esta predicción. Pero hemos realizado experimentos equivalentes con partículas elementales, y el efecto es incontrovertiblemente real. La cantidad de tiempo experimentada por un viajero depende del camino que tome a través del espacio-tiempo, exactamente de la misma manera que la distancia recorrida a través del espacio depende del camino que siga por él.

Como hemos señalado, nada de esto es paradójico. Pero sí supone un cambio radical en el modo de pensar acerca de la naturaleza del tiempo. En el universo de Newton, el tiempo era universal, y los relojes no iban a perder la sincronización solo porque alguien se diera un rápido paseo por el espacio. En el universo de Einstein, el tiempo es personal, y su reloj reflejará la particularidad de su viaje.

LA VELOCIDAD DEL TIEMPO

El experimento mental de los gemelos ha generado una gran cantidad de discusiones, no siempre útiles, en torno al «ritmo con que pasa el tiempo». Si Bob no ha envejecido tanto como Alice, es sin duda tentador decir que el reloj de Bob iba más lento. Ceder a esta tentación conduce a todo tipo de confusiones.

Bajo determinadas circunstancias —cuando estamos nerviosos, o aburridos, o expectantes por algo, o puramente aterrados—, puede parecer que la cadencia con que pasa el tiempo cambia. Nos parecerá que el tiempo va más deprisa, o más lento, dependiendo de nuestra situación. Puede quedarse tranquilo, este sentimiento es una cuestión enteramente biológica y psicológica, no una cuestión de física. Su reloj interno, artilugio biológico

poco fiable, se ha desajustado. El tiempo en sí mismo no puede ir más deprisa o más despacio.

Porque, a fin de cuentas, ¿qué significa que el tiempo vaya «más deprisa»? La velocidad es la razón a la que algo sucede (¡la derivada!) con respecto al tiempo. La velocidad a la que usted viaja es la razón a la que va cubriendo distancia a medida que pasa el tiempo. La razón a la que se llena un contenedor es el volumen del líquido introducido con relación al tiempo. Así pues, la velocidad del tiempo, si tal concepto tuviera algún sentido, sería la razón a la que el tiempo pasa con respecto al tiempo. Esto es: un segundo por segundo. Literalmente, no puede ser otra cosa. Tendrá que ser una cantidad cuyo sentido deberá preocuparnos tan poco como si nos preguntáramos por el número de metros que recorremos por metro recorrido.

Oirá decir a veces que, según la teoría de la relatividad, el tiempo puede acelerarse o ralentizarse. Es un disparate. O por decirlo de un modo más educado: es un modo desorientador de describir un fenómeno real.

La nueva característica descrita por la relatividad es que la duración total experimentada por dos observadores que se desplazan de maneras diferentes, generalmente no será igual, aunque comiencen y finalicen en los mismos sucesos del espacio-tiempo. Esto no se debe a que la razón del tiempo varíe, sino a que determinada persona se ha desplazado siguiendo una trayectoria diferente y, por tanto, ha cubierto una cantidad diferente de espacio-tiempo. Si una persona camina en línea recta entre dos puntos, y otra sigue un camino sinuoso entre esos mismos puntos, la segunda persona recorre una distancia diferente, pero no diremos que han experimentado un número diferente de metros por metro.

Cuando hay personas que dicen que el tiempo va más despacio, de acuerdo con la teoría de la relatividad, suele ser porque no se han liberado, verdaderamente, hasta el núcleo de su ser, de la noción newtoniana de tiempo absoluto. De un modo implícito, están relacionando el tiempo que transcurre en el reloj de Bob con un tiempo objetivo existente allá fuera en el universo. Pero no existe tal tiempo objetivo o absoluto: solo existe aquello que mide el reloj. Podemos perfectamente establecer una coordenada temporal en el espacio-tiempo, pero eso dista mucho de ser el

tiempo objetivo. Yo podría usar un sistema de coordenadas diferente del suyo, y no hay nada en el mundo físico que dependa de tal opción.

Mientras Bob va a toda velocidad en su nave espacial, si en algún momento se le ocurre mirar el reloj, verá que avanza a su ritmo normal de un segundo por segundo. A fin de cuentas, tanto Bob como el reloj están desplazándose siguiendo la misma trayectoria a través del espacio-tiempo y, por tanto, experimentan la misma cantidad de tiempo propio. Si Bob se toma el pulso, obtendrá como resultado las mismas pulsaciones por minuto que sean habituales en él. (Al margen de la excitación que pueda sentir por estar en una nave a tan enorme velocidad, lo cual sería comprensible).

Pero… quizás esté pensando —y este es un punto crucial—, seguramente el tiempo de Bob se haya ralentizado con respecto al tiempo de Alice, ¿o no? Porque, después de todo, cuando vuelvan a reunirse, él habrá envejecido menos.

El problema está en que, si queremos demostrar que tiene sentido la afirmación de que el reloj de Bob avanza a ritmo diferente del de Alice, tendríamos que compararlos. Esto no es ningún problema cuando los relojes están en el mismo lugar y podemos observarlos sin más. Pero cuando están separados, no hay forma de comprobar «simultáneamente» qué hora marca cada reloj. Si estamos junto a uno de los relojes, pero lejos del otro, cualquier tipo de señal procedente del reloj alejado tardaría un tiempo en llegar hasta nosotros; ninguna señal que tuviera que partir de allí para llegar hasta aquí podría desplazarse a mayor velocidad que la de la luz. Cuando pudiéramos observar lo que señala el reloj alejado, habría pasado ya cierto tiempo en el reloj cercano.

Tal vez piense que esto no es más que un contratiempo técnico, que podríamos solventar haciendo saltar señales de un reloj a otro, o por algún otro método por el estilo. Cualquier estrategia que pueda ocurrírsele será bienvenida. Pero la comparación jamás será absoluta o instantánea, no hay forma. Habrá ideado algo arbitrario, no un medio natural de comparar relojes distantes entre sí.

La estrategia correcta pasa por renunciar a la idea de comparar relojes que están a gran distancia uno de otro. Es lo mejor, y está en el espíritu de la

relatividad. Piense en términos de lugar y abandone el gazmoño impulso de extender sus pensamientos reduccionistas hasta convertirlos en estructuras absolutas aplicadas a todo lo que abarca el espacio y el tiempo.

EL ESPACIO-TIEMPO DE MINKOWSKI

Ha tenido mucha paciencia con tanta palabra. Es hora de volver a las ecuaciones.

Hemos insistido en que el tiempo medido en relojes en movimiento a través del espacio-tiempo es análogo a la distancia recorrida a lo largo de una trayectoria por el espacio. Podemos dar un poco de contenido a todo esto si pensamos en el modo en que medimos las distancias.

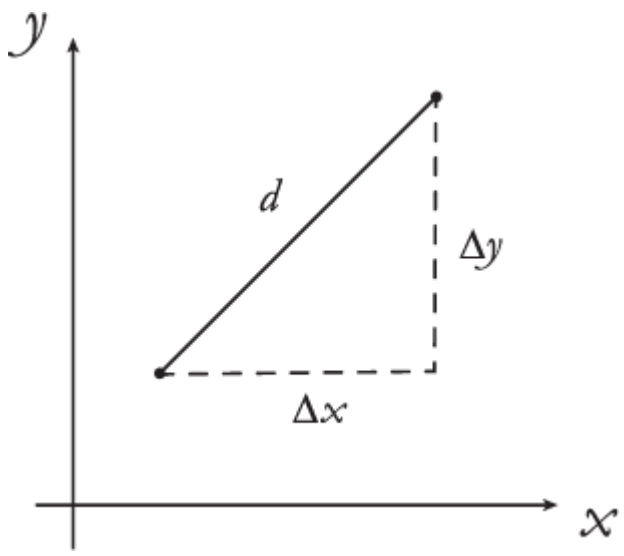
La repuesta estriba en el teorema de Pitágoras. Consideremos un triángulo rectángulo, uno de cuyos ángulos interiores es ángulo recto. El lado más largo, la hipotenusa, será el lado opuesto al ángulo recto. Pitágoras nos dice que el cuadrado de la longitud de la hipotenusa es igual a la suma de los cuadrados de las longitudes de los otros dos lados: $a^2 + b^2 = c^2$, siendo a y b los lados más cortos y c la hipotenusa.

Este resultado es interesante por sí mismo, sobre todo para los fans de los triángulos rectos. Pero cuando se vuelve fundamental para nosotros es cuando expresamos el espacio en forma de coordenadas cartesianas (perpendiculares). Para mayor simplicidad, consideraremos un espacio bidimensional y designaremos a nuestras coordenadas como x e y, como es habitual.

Para dos puntos cualesquiera en el plano, se da una distancia d bien definida entre ambos. De acuerdo con nuestro sistema de coordenadas, hay una forma precisa de calcular cuál es esa distancia. Podemos dibujar un triángulo rectángulo partiendo de un punto, desplazándonos en la dirección x (horizontalmente) hasta alcanzar el valor de x del segundo punto, y desplazándonos entonces en la dirección y (verticalmente) hasta alcanzar el

segundo punto propiamente dicho. Queda así definido un triángulo rectángulo, cuyos lados menores tienen por longitud x y y, y cuya hipotenusa es la distancia que buscamos. (Recuerde que la letra griega , delta, significa «cambio en»). Así, aplicando el teorema de Pitágoras, la distancia puede expresarse en los términos de los cambios en las coordenadas entre los dos puntos, del modo siguiente:

$$d^2 = (\Delta x)^2 + (\Delta y)^2 \qquad (6.1)$$



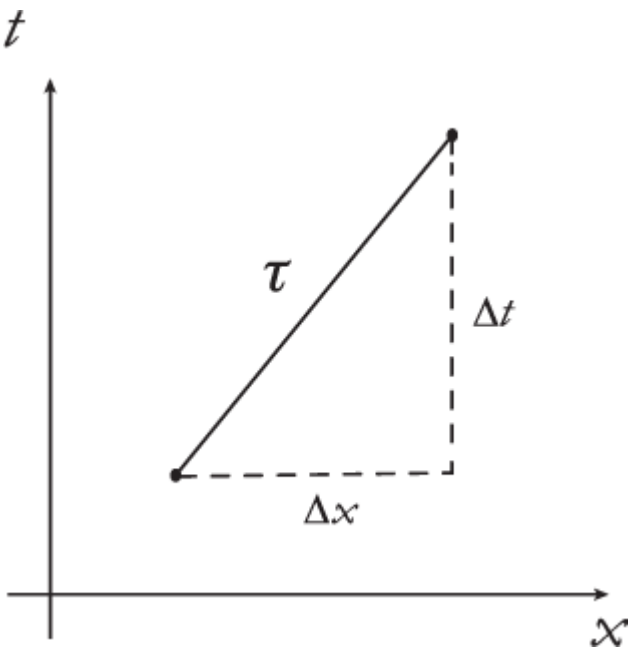
Probablemente, usted ya sabía todo esto. Lo bueno es que casi lo mismo funciona para el tiempo medido a lo largo de una trayectoria en el espacio-tiempo. «Casi», porque es preciso hacer una modificación crucial: en el espacio-tiempo, en el teorema de Pitágoras se introduce un signo menos. Este signo menos es el responsable del paso de «la distancia más corta» a «el tiempo más largo» en relación con las trayectorias rectilíneas.

Considere un espacio simplificado de dos dimensiones, en que hemos definido las coordenadas x para el espacio e t para el tiempo. E imagine que hay una trayectoria rectilínea (una velocidad constante) que va de un suceso a otro suceso. Sea τ (la letra griega tau) el tiempo propio, aquel que un reloj mediría en el transcurso del recorrido. Una vez más, podemos definir las

diferencias en las coordenadas entre los dos sucesos, en esta ocasión Δx para el espacio y Δt para el tiempo.

La característica definitoria del espacio-tiempo de Minkowski, es escenario en el que se representa el drama de la relatividad especial, es que el tiempo propio cumple con una ecuación con la forma del teorema de Pitágoras, pero con un signo menos crucial delante del término referente al espacio:

$$\tau^2 = (\Delta t)^2 - (\Delta x)^2 \qquad (6.2)$$



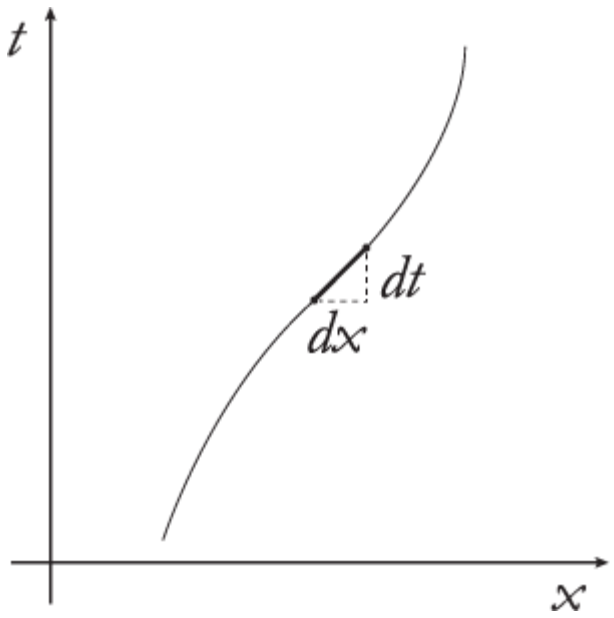
Esta simple ecuación nos dice mucho acerca de cómo funciona el espacio-tiempo (todo cuanto necesitamos saber, en realidad). Piense en un observador estacionario, una persona que permanece inmóvil en el sistema de coordenadas que hemos establecido. Envejecerá en el tiempo, por supuesto, pero no experimenta cambios en el espacio. De modo que, para un observador estacionario, tenemos que $\tau^2 = (\Delta t)^2$, o simplemente $\tau = \Delta t$. El tiempo propio y el tiempo de las coordenadas son el mismo, que es a lo que la mayoría de nosotros estamos acostumbrados, por lo demás.

Los observadores que se mueven cuentan otra historia. Tendrán un Δx diferente de cero. Así que su tiempo propio siempre será menor que el de un observador estacionario, dada una cantidad fija de tiempo transcurrido en las coordenadas, y todo a causa de ese divertido signo menos de (6.2). La relatividad plantea una especie de compensación entre el tiempo y el espacio: en un desplazamiento entre dos sucesos cualesquiera fijados de antemano, cuanto más nos movemos en el espacio, menos tiempo transcurrido experimentamos.

Hemos estado hablando acerca de trayectorias rectilíneas, pero no hay ningún problema en absoluto en incorporar recorridos arbitrarios en nuestra historia. Podemos imaginar cuál sería la estrategia: aplicar una versión infinitesimal de (6.2) a cada uno de los cortísimos segmentos de una trayectoria, y luego recurrir al cálculo. Para distancias infinitesimales en el espacio-tiempo, la fórmula modificada de Pitágoras se convierte en:

$$d\tau^2 = dt^2 - dx^2 \qquad (6.3)$$

Si queremos calcular la cantidad total de tiempo propio transcurrido durante una trayectoria, solo tenemos que buscar la integral: $\Delta\tau = \int d\tau$.



Hay personas que sugieren a veces que la relatividad especial funciona tan solo para trayectorias carentes de aceleración, y que se necesita la relatividad general para tratar la aceleración. Tonterías. La relatividad general cobra mayor importancia cuando el espacio-tiempo es curvo y entra en juego la gravedad. Mientras el espacio-tiempo sea plano —como en el caso del espacio-tiempo de Minkowski, del que no nos estamos apartando en este capítulo—, es aplicable la relatividad especial, y pude considerar cualquier tipo de trayectoria que desee.

LA VELOCIDAD DE LA LUZ

Es el momento de hacer una confesión. Hemos hecho una pequeña trampa en las relaciones pitagóricas, en las ecuaciones (6.2) y (6.3). Recuerde el análisis dimensional: las cantidades físicas se expresan en unidades, y solo podemos sumar cantidades que estén expresadas en las mismas unidades. Pero τ^2 y $(\Delta t)^2$ se miden en unidades de (tiempo)^2, mientras que $(\Delta x)^2$ se mide en unidades de (distancia)^2. ¿Quiénes nos hemos creído que somos, para sumar cantidades de esta manera?

El secreto está en que la relatividad proporciona un factor de conversión universal entre el espacio y el tiempo. (Ambos son direcciones en el espacio-tiempo, al fin y al cabo). Este factor de conversión se representa con la letra c, y es numéricamente igual a:

$$c = 299{,}792{,}458 \text{ metros/segundo} \qquad (6.4)$$

Es posible que conozca esta cifra como la velocidad de la luz. Pero lo importante no es que la luz viaje a esa velocidad, sino el hecho de que haya una velocidad universal, imbricada en el tejido mismo del espacio-tiempo, que podemos utilizar como conversor entre espacio y tiempo. Simplemente

sucede así: las ondas de luz (al igual que las ondas gravitatorias, por cierto) viajan a esa velocidad tan especial.

Es una propiedad tan importante, que unidades como los «metros» quedan anticuadas por incómodas. Es enormemente más funcional utilizar unidades como «segundos luz», la distancia que la luz recorre en un segundo. Es decir, un segundo luz equivale a 299.792.458 metros. Esta cifra es exacta, porque es más fácil definir un segundo en términos de ciertas vibraciones atómicas, que definir el metro como la distancia que recorre la luz en 1/299.792.458 segundos.

Puede que sea exacto, pero es un número muy incómodo. Si utilizamos segundos luz en lugar de metros, podemos escribir sencillamente:

$$c = 1 \text{ segundo luz/segundo} \qquad (6.5)$$

Es igual de exacto, y mucho más fácil de recordar. El resto funciona igual de bien si utilizamos años luz y años, etcétera.

Lo bueno de utilizar unidades en que c= 1, es que podemos omitir sin más la velocidad de la luz de nuestras ecuaciones, ya que multiplicar o dividir por 1 no afecta a las mismas. Esto es lo que subrepticiamente hicimos al anotar (6.2). En la parte de la derecha, donde escribimos Δx, lo que realmente queríamos decir era $\Delta x/c$. Una vez hemos optado por unidades en que c = 1, podemos considerar la velocidad de la luz implícita. Además de ser más limpio desde el punto de vista de la expresión, ayuda a subrayar una verdad profunda: que las distancias y los tiempos son medidas ambas de desplazamiento en el espacio-tiempo.

Podemos ahora atribuir números al experimento mental de los gemelos. Consideremos el segmento del viaje de Bob desde el punto de partida hasta el punto en que da media vuelta. (El segundo segmento es una repetición del primero, solo que con la dirección del viaje invertida, así que no necesitamos realizar el cálculo dos veces). Asumimos que se desplaza a velocidad constante:

$$v = \frac{\Delta x}{\Delta t} \qquad (6.6)$$

De forma equivalente, sabemos que $\Delta x = v\Delta t$. Haciendo las sustituciones correspondientes en (6.2):

$$\begin{aligned} \tau^2 &= (\Delta t)^2 - (\Delta x)^2 \\ &= (\Delta t)^2 - v^2 (\Delta t)^2 \\ &= \left(1 - v^2\right)(\Delta t)^2 \end{aligned} \qquad (6.7)$$

o:

$$\tau = \sqrt{1 - v^2}\,\Delta t \qquad (6.8)$$

Alice y Bob comienzan y terminan en los mismos sucesos, y atraviesan la misma cantidad de tiempo de coordenadas . Alice permanece estacionaria, de modo que su tiempo propio es igual al tiempo de coordenadas transcurrido. Pero Bob experimenta menos tiempo propio, por una cantidad multiplicativa $\sqrt{1 - v^2}$ (en unidades en que c = 1, recuerde). Si v = 0,99, tenemos que $\sqrt{1 - v^2} = 0{,}14$, y 1/0,14 es aproximadamente 7. Por eso Alice envejece siete años por cada año que envejece Bob.

Este cálculo explica también por qué se tardó tanto tiempo en postular la relatividad. Newton era inteligente, ¿por qué llegó a pensar que podíamos incluir nuestro tiempo propio en un tiempo absoluto? La respuesta es que la mayoría de nosotros vivimos toda nuestra vida moviéndonos a una velocidad mucho más lenta que la de la luz. Si conducimos un coche a 100 km/h, tenemos que $v = 10{-}7$ en unidades en que c = 1. En tal caso, el factor

$\sqrt{1-v^2}$ de (6.8) es aproximadamente 0,999999999999995. Es una cifra tan próxima a uno, que para nosotros la diferencia es absolutamente imperceptible. En nuestra vida diaria, los tiempos propios que experimentamos son virtualmente indistinguibles de un marco simple de coordenadas temporales.

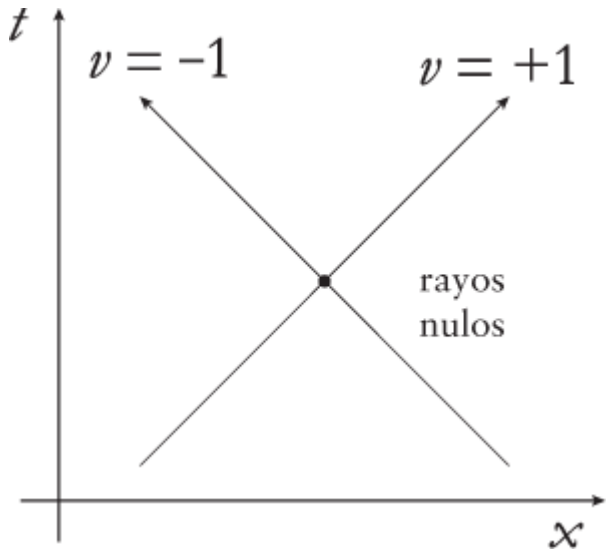
CONOS DE LUZ

El signo menos en la fórmula del tiempo propio (6.2) abre una interesante posibilidad. Si consideramos una trayectoria rectilínea que atraviesa cantidades iguales de espacio y tiempo, $(\Delta x)^2 = (\Delta t)^2$, tendremos $\tau = 0$. El objeto se mueve, pero en su desplazamiento no transcurre tiempo alguno. Son simplemente trayectorias que se desplazan a la velocidad de la luz:

$$v = \frac{\Delta x}{\Delta t} = \pm 1 \qquad (6.9)$$

(Este signo menos solo significa que el objeto podría desplazarse a la izquierda en lugar de hacerlo a la derecha, lo cual no importa para nuestro propósito actual).

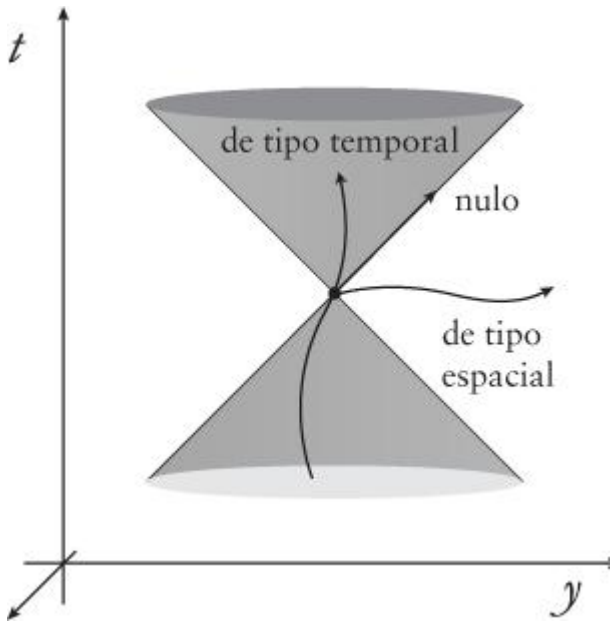
Es un resultado muy interesante. Cualquier cosa que se desplace a la velocidad de la luz, incluida la luz misma, no experimenta el paso del tiempo. Es tentador preguntarse cómo sería viajar a la velocidad de la luz. (Al parecer Einstein se preguntaba esto mismo cuando iba al instituto). La respuesta más rápida es que no sería «como nada» viajar a la velocidad de la luz. Si hubiera algún modo de hacerlo, no experimentaríamos el paso del tiempo, así que no tendríamos ningún tipo de percepción ni de pensamiento consciente.



Las trayectorias espacio-temporales que se desplazan a la velocidad de la luz se llaman trayectorias nulas (porque el tiempo propio es cero), o trayectorias cercanas a la velocidad de la luz (por razones obvias). Cuando elegimos unidades en que c = 1, las trayectorias nulas se representan en forma de líneas diagonales con una inclinación de 45 grados en un diagrama de espacio-tiempo, y atravesando cantidades iguales de espacio y tiempo. Cualquier cosa que se mueva a velocidad inferior a la de la luz atravesará más tiempo que espacio, por lo que tales trayectorias se conocen como trayectorias de tipo temporal, y se desplazan hacia arriba en el diagrama de espacio-tiempo. Podemos considerar también las trayectorias de tipo espacial, que se mueven hacia los lados y atraviesan más espacio que tiempo. No hay ningún objeto físico capaz de viajar siguiendo trayectorias de tipo espacial (no podemos ir a una velocidad superior a la de la luz), pero ello no nos impide representarlas gráficamente o pensar en ellas.

Tomemos un suceso simple en el espacio-tiempo y llamémoslo A. Imaginemos que dibujamos todos los rayos similares a los de la luz, partiendo de ese punto y expandiéndose hacia el futuro. Es como si hubiéramos encendido una bombilla y la hubiéramos apagado inmediatamente en tal punto, y trazáramos las trayectorias de todos los fotones de luz creados en ese instante. (Como una versión espacio-temporal de meter el dedo en un remanso de agua y ver cómo se expande una onda

sobre la superficie, si tal onda se desplazara a la velocidad de la luz). Agrupadas, todas esas trayectorias constituyen el cono de luz del suceso A. Nuestra bombilla hipotética no tiene por qué existir: todo suceso define un cono de luz de rayos nulos asociado a él, aunque no haya luz alguna viajando a través de esos rayos. En realidad, existe un cono de luz pasado (como también un cono de luz futuro), consistente en todas las trayectorias similares a los rayos de luz que estarían apuntando directamente hacia A desde el pasado.

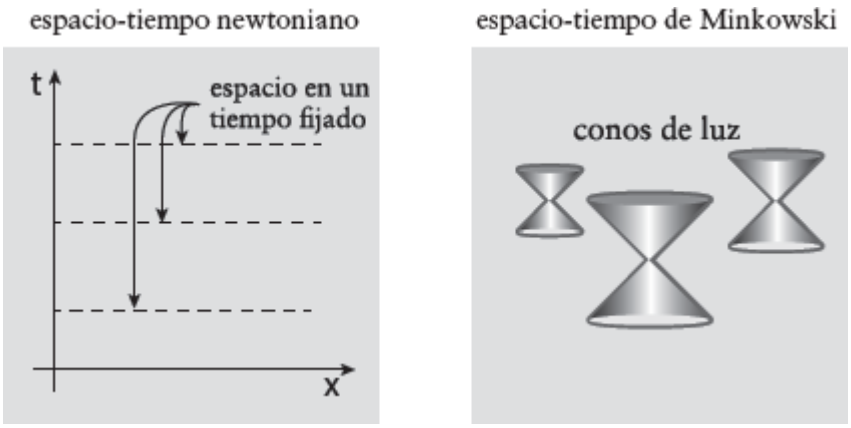


El límite de la velocidad de la luz implica que las trayectorias cuya existencia es posible de acuerdo con las leyes de la física, que atraviesan un suceso en su paso del pasado al futuro, tienen que permanecer dentro de los límites de los conos de luz de dicho suceso. Esto vale para todo suceso. Cada punto en el espacio-tiempo define conos de luz pasados y futuros, y cualquier trayectoria física que pase a través de ellos debe permanecer dentro de sus límites.

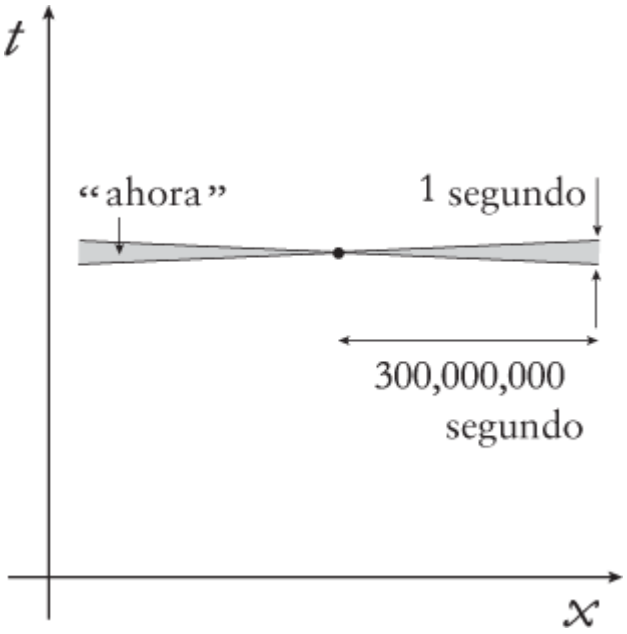
Esta estructura de tipo cono de luz es lo que sustituye a la curiosa noción newtoniana del espacio y tiempo absolutos. Cuando dibujamos un diagrama espacio-temporal newtoniano, escindimos el espacio-tiempo en «momentos

de tiempo» horizontales. Dos sucesos que están separados en el espacio, pero ocurren al mismo tiempo, se dice que son «simultáneos».

No es esta la manera correcta de pensar una vez hemos hecho la transición a la relatividad. No existe noción alguna de simultaneidad entre dos sucesos separados. Solo existe la posibilidad de que dos sucesos estén, cada uno de ellos, dentro de los respectivos conos de luz del otro. Si lo están, decimos que los sucesos están «al modo del tiempo»; mientras que si están fuera, están separados «al modo del espacio». Tales nociones están bien definidas, y sobre ellas estará de acuerdo todo observador en el universo; lo cual no es el caso de «simultáneo».



Escindir el espacio-tiempo en momentos horizontales parece algo natural porque, una vez más, en nuestra vida diaria nos movemos muy lentamente en comparación con la velocidad de la luz. Para los fines cotidianos, es mucho más adecuado utilizar unidades como «metros» que otras como «segundos luz». (En toda la historia de la humanidad, únicamente unos pocos astronautas del programa Apolo han llegado a trasladarse más allá de un segundo luz desde su lugar de nacimiento en la Tierra, y hacerlo les costó mucho más tiempo que un segundo). Si dibujamos un diagrama espacio-temporal utilizando metros y segundos, los conos de luz no presentan una inclinación de 45 grados. Atraviesan tantos metros en un solo segundo, que parecen casi horizontales. Es difícil decir cuál es la diferencia entre «el conjunto de todos los sucesos separados al modo del espacio» y «un solo momento de tiempo». Pero esa diferencia es muy real.



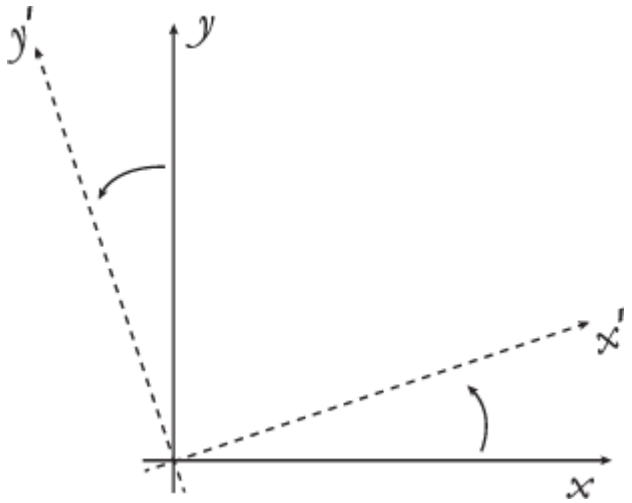
Una vez se haya convertido en un acérrimo de la relatividad, haría bien en no dibujar «cortes de espacio en un tiempo fijado» en sus diagramas espacio-temporales. Debería dibujar únicamente conos de luz, que son objetivos y universales. El impulso de escindir el espacio-tiempo en momentos de tiempo es casi irresistible. Simplemente, no piense que esté reflejando ningún tipo de característica verídica de la realidad.

MARCOS DE REFERENCIA

Dicho esto, aceptemos que las personas van a seguir trazando cortes horizontales en sus diagramas espacio-temporales, aunque hayan aprendido que hay mejores representaciones. Tales cortes se llaman «marcos de referencia», o «marcos de referencia globales», para subrayar que los extendemos a todo el espacio. Cuando realizamos la transición desde el espacio-tiempo plano de la relatividad especial hasta el espacio-tiempo curvo de la relatividad general, los marcos de referencia globales de ser «innecesarios» a «quizá ni tan solo posibles». Pero pensar en ellos nos

ayuda a entender por qué las personas hablan de ideas como «contracción de la longitud», de modo que es mejor dar explicación de ellos, aunque solo sea para comunicarnos mejor con nuestros amigos entusiastas de la relatividad especial.

Comencemos, como es habitual, recordando cómo funcionan las cosas en el espacio. En un espacio plano de dos dimensiones, a menudo es conveniente utilizar coordenadas cartesianas x e y, que definen ejes perpendiculares entre sí. Pero todos sabemos que lo que cuenta es el plano bidimensional mismo, no las coordenadas particulares que utilicemos para señalar puntos en él. Podríamos elegir otras coordenadas x' e y', por ejemplo, haciendo rotar las originales en un ángulo determinado. Las nuevas coordenadas serán igual de útiles que las viejas, y las cantidades físicamente mensurables como la distancia entre dos puntos no cambiarán.

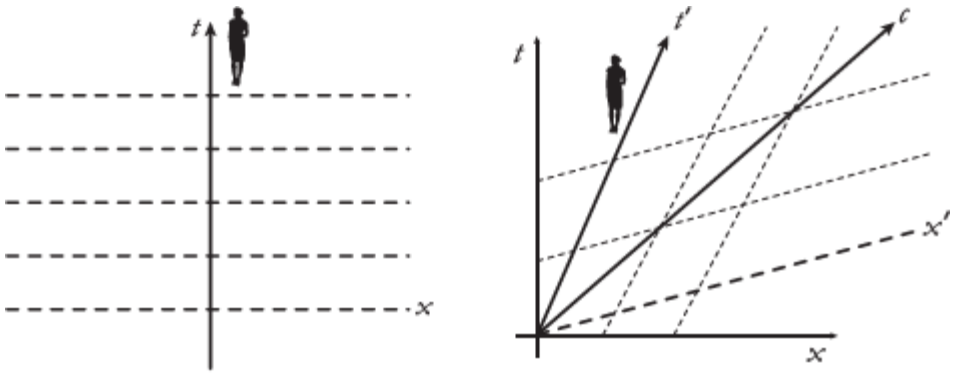


Podemos hacer lo mismo en el espacio-tiempo. Consideremos un único observador que no experimenta ningún tipo de aceleración y las utiliza para establecer un sistema de coordenadas. Llevan un reloj consigo, por lo que podemos utilizar el tiempo propio de su trayectoria para definir una coordenada temporal t. Además, aplicando un poco más de creatividad, imaginaremos que cada segundo nuestro observador dispara rayos «infinitamente veloces» en todas direcciones. Todos sabemos que no hay nada que pueda viajar infinitamente deprisa, así que tal desplazamiento es puramente conceptual. Lo que queremos significar es que nuestro

observador define líneas espaciales que son «perpendiculares» a su línea de universo. Llamaremos x a la distancia longitudinal de estos rayos. De este modo, definimos un sistema de coordenadas —un marco de referencia inercial, puesto que nuestro observador no experimenta aceleración y es, por tanto, «inercial»— que se extiende por el espacio-tiempo.

Probablemente habrá adivinado lo que viene a continuación: haremos lo mismo, partiendo ahora de un observador diferente, que tampoco experimenta aceleración, pero que se mueve a velocidad constante con respecto al primero. Definen una coordenada temporal t' utilizando su reloj, y disparan rayos espaciales imaginarios de velocidad infinita para definir una coordenada espacial x'. Basar un nuevo sistema de coordenadas en un observador en movimiento es análogo a rotar un sistema cartesiano de coordenadas en un espacio ordinario. Este cambio de coordenadas se llama transformación de Lorentz, por el físico holandés Hendrik Antoon Lorentz.[1]

Pero hay una sorpresa. Cuando trazamos los rayos de velocidad infinita emitidos por el observador en movimiento a nuestro diagrama espacio-temporal original, no parecen perpendiculares al movimiento del observador. El nuevo «eje temporal» y el nuevo «eje espacial» (t', x') parece como si se entrecortaran entre sí, en lugar de conservar sus ángulos rectos. Pero en realidad son perpendiculares entre sí; la diferencia puede rastrearse hasta aquel signo menos de la fórmula de la distancia de Minkowski (6.2).



marco del observador originario

marco del observador en movimiento

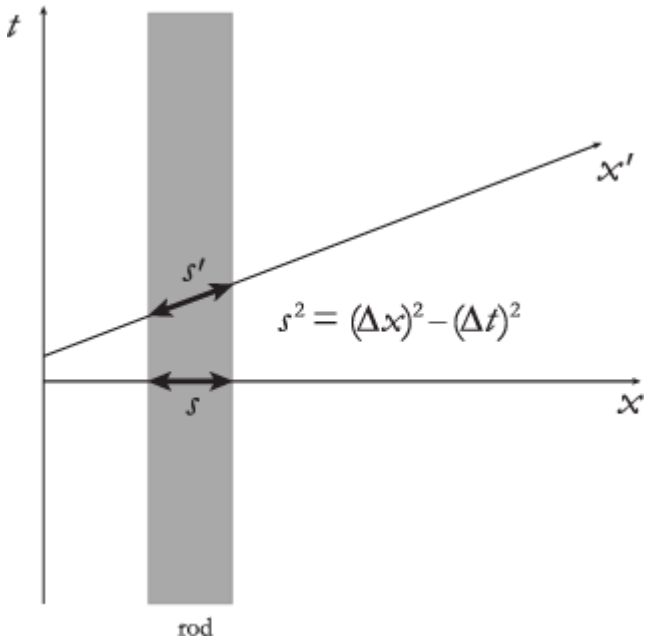
Desde una perspectiva física, este efecto de entrecruzamiento puede entenderse como consecuencia de la constancia de la velocidad de la luz. Advertimos que los nuevos ejes del tiempo y del espacio mantienen un ángulo constante con respecto a los conos de luz. Ello es reflejo del hecho de que todo observador medirá la velocidad de la luz con igual resultado, sea cual sea el marco referencial que utilicen.

Es este un buen lugar para mencionar que la expresión «teoría de la relatividad» resulta inapropiada. El significado básico de «relatividad» en este contexto es que no existe en el universo un marco de referencia objetivo, privilegiado; la velocidad de un objeto solo podemos medirla en relación con el movimiento de otros objetos, y no en un sentido absoluto. Pero esto también era así en la mecánica newtoniana, que asumía la relatividad galileana. La «teoría de la relatividad» en sentido moderno proviene de la combinación de este principio de relatividad con la constancia de la velocidad de la luz para cualquier observador. O dicho de un modo igual de correcto y más elegante, proviene de la idea de que vivimos en el espacio-tiempo de Minkowski, en que el tiempo propio se mide por (6.2).

CONTRACCIÓN DE LA LONGITUD

La inclinación de los marcos de referencia ayuda a explicar el famoso fenómeno de la contracción de la longitud, según el cual los objetos que se mueven a alta velocidad reducirían su longitud. Al fin y al cabo, ¿qué es la «longitud» de algo? Usted podría pensar que una regla, por ejemplo, tiene determinada longitud. Pero, a menos que su regla existiera literalmente tan sol por un instante, posee extensión temporal, además de espacial. Si simplificamos nuestras vidas reduciendo una regla física a un objeto con tan solo una dimensión de extensión espacial, seguirá teniendo un «volumen de universo» bidimensional en el espacio-tiempo. Cuando hablamos de su longitud, necesitamos elegir un marco de referencia para distinguir la parte espacial de la parte temporal del volumen de universo de la regla. La «longitud de la regla» es en realidad la distancia espacial de una sección

transversal de ese volumen de universo en un marco de referencia particular. Habitualmente, con buen criterio, lo que tenemos en cuenta es el marco en reposo de la regla, el sistema de coordenadas en el cual permanece inmóvil. Hablar de la longitud de una regla en movimiento es equivalente a medir esa distancia espacial en un marco diferente.



En el diagrama podemos ver cómo ciertamente parece que la sección transversal espacial de nuestra regla será diferente en un marco de referencia en movimiento. Por desgracia, en el marco referencia en movimiento parece más larga, en lugar de más corta. Diríase que hay elongación, más que contracción. ¿Qué está pasando?

La culpa la tiene nuestra intuición acerca de cómo funciona la longitud cuando trazamos una representación del espacio, que no siempre puede trasponerse a diagramas espacio-temporales. La fórmula (6.2) nos dice cómo calcular el tiempo propio de una trayectoria en términos de desplazamiento en las coordenadas de tiempo y de espacio. Pero «tiempo propio» solo tiene sentido para trayectorias de tipo temporal. Para trayectorias de tipo espacial, Δx es mayor que Δt, lo cual significa que t^2 sería negativa, cualidad que no es pertinente para el cuadrado de un número. (No tiene sentido pensar en un intervalo espacio-temporal como un

número imaginario). No hay razón para atribuir tiempo propio a un intervalo de tipo espacial.

Lo que queremos atribuir a los intervalos de tipo espacial es, por supuesto, una distancia espacial, que en este contexto solemos representar con la letra s. Por fortuna, hay una versión de (6.2) que funciona para este propósito: solo hay que invertir el signo. Esta es la fórmula para la longitud de un segmento de tipo espacial:

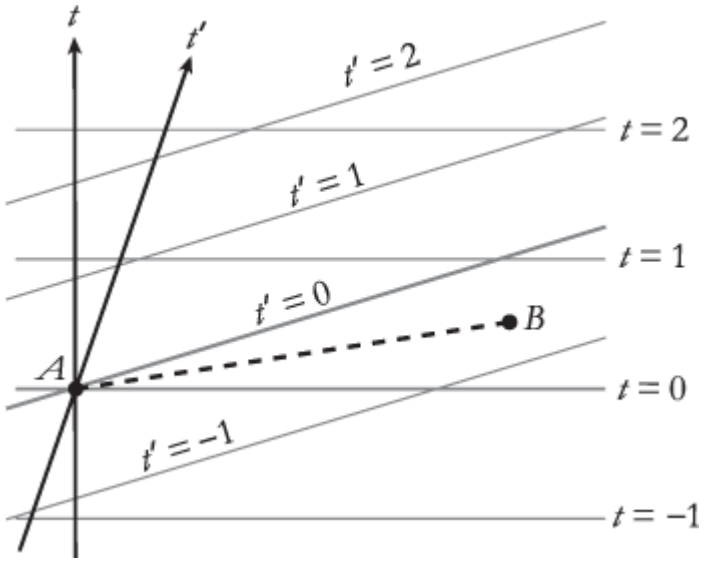
$$s^2 = (\Delta x)^2 - (\Delta t) \qquad (6.10)$$

La diferencia de ese signo menos entre el tiempo y el espacio explica por qué la longitud de la regla medida en el sistema en movimiento es realmente más corta que en el sistema en reposo, aunque parezca más larga en el diagrama: porque se extiende tanto en el tiempo como en el espacio. En un plano espacial bidimensional, al coger el extremo de un segmento lineal horizontal y desplazarlo verticalmente, el segmento siempre se alarga, de acuerdo con el teorema de Pitágoras. Pero en el espacio-tiempo, al mover un poco el extremo de un segmento en el tiempo, el segmento se acorta. La contracción de la longitud es real, pero no significa que el objeto se encoja físicamente; es solo que estamos en un marco de referencia diferente, midiendo una cantidad diferente.

LA SIMULTANEIDAD Y SUS SINSABORES

Al comparar estos dos marcos de referencia, volvemos a la dura realidad de que, en la relatividad, no existen sucesos distantes que se den «al mismo tiempo». Así se desprende con bastante obviedad del diagrama, puesto que las líneas definitorias de «simultáneo» no son las mismas en los dos sistemas de coordenadas. Pero la cosa aún puede ir a peor.

Tomemos el suceso en que nuestros dos observadores se solapaban (punto de origen de ambos sistemas de coordenadas) y llamémoslo A. Designemos luego otro suceso B, alejado de A en el espacio, y tan solo un poco en el tiempo futuro, de acuerdo con la medición en nuestra coordenada temporal originaria t.



Tal como se desprende del diagrama, vemos de inmediato que, si bien B está situado en el futuro de A en el sistema de coordenadas (x, t), está en el pasado de A en el sistema de coordenadas (x', t').

Así es la vida para los sucesos separados de acuerdo con el modelo espacial. No es pertinente preguntarse qué suceso está «realmente» en el pasado o en el futuro. Eso depende del marco de referencia que utilicemos, y todos los marcos de referencia son igualmente válidos. Todo lo que podemos decir, en buena conciencia, es que los dos eventos están separados de acuerdo con el tipo de separación espacial.

Para nosotros es natural preguntarse qué está sucediendo «ahora mismo» en cualquier otro lugar del mundo. Pero esta forma de pensar nos resultará problemática cuando nos pongamos a considerar distancias de magnitud astronómica. La estrella más cercana nuestro sol, Próxima Centauri, se encuentra a una distancia aproximada de cuatro años luz. Para cualquier

suceso particular de la Tierra, hay un abanico de ocho años de sucesos en Próxima Centauri que podrían considerarse «simultáneos» a tal suceso terrestre, dependiendo de nuestro marco de referencia.

Esta característica de la relatividad especial representa un problema para Star Trek y otras space operas. Es habitual que en estas historias se proponga la invención del desplazamiento por curvatura o de cualquier otro adelanto tecnológico que permite a nuestros héroes viajar a una velocidad superior a la de la luz. Pero una vez consigue uno viajar siguiendo trayectorias de tipo espacial, puede efectivamente también viajar hacia atrás en el tiempo, al menos en algún marco de referencia. De hecho, todas las trayectorias de tipo espacial están creadas igual, de acuerdo con la relatividad especial. De modo que si uno puede viajar por encima de la velocidad de la luz, también puede retroceder en el tiempo en cualquier marco de referencia.

Quizás esto sea lo que desee para su universo de ciencia ficción. (Star Trek ha mostrado bastante desvergüenza al recurrir a los viajes en el tiempo cuando la cosa parecía divertida, e ignorarlos luego en los episodios posteriores). Aunque bien es verdad que es difícil mantener una lógica consistente en estas cuestiones. Es más prudente, tanto para la física como para la ficción, excluir enteramente los viajes a velocidad superior a la de la luz.

UNIFICACIÓN

La teoría del electromagnetismo de James Clerk Maxwell fue uno de los primeros grandes ejemplos de unificación dentro de la física, por la que se agruparon los fenómenos eléctricos y los magnéticos bajo un mismo título. (Un ejemplo anterior aún es el de la ley de gravitación universal de Newton, que unificó la caída de las manzanas con el movimiento de los planetas). Aportó numerosas ideas aparejadas a ella, como noción de que la luz es una forma de onda electromagnética. La relatividad especial, que fue inspirada por la teoría de Maxwell, es otro gran ejemplo, al unificar el espacio y el

tiempo en espacio-tiempo. No es de sorprender que esta perspectiva trajera implícitas sus propias nuevas ideas.

En el espacio tridimensional, a menudo hablamos de vectores, como el vector velocidad de un objeto en movimiento. Como sabemos, no es más que la derivada de la posición del objeto con respecto al tiempo: . Dado un sistema de coordenadas en el espacio, es conveniente expresar el vector en términos de sus componentes en cada dirección:

$$v^i = \left(v^x, v^y, v^z\right) = \left(v^1, v^2, v^3\right) \qquad (6.11)$$

Aquí, la i volada es un índice —una marca para distinguir de qué componente estamos hablando, ¡no un exponente!— que sigue un orden numérico (1, 2, 3) y que se corresponden con las direcciones espaciales (x, y, z).[2] Advierta la flexibilidad de esta notación: la i puede representar tanto el índice numérico del componente que estemos considerando, como la variable literal de las coordenadas.

Ahora que hemos avanzado y pasado ya del espacio al espacio-tiempo, podemos con naturalidad definir una generalización de estos vectores tridimensionales, a los que imaginativamente llamaremos cuadrivectores, ya que el espacio-tiempo es cuadridimensional. La versión espacio-temporal de la velocidad es la cuadrivelocidad, que podemos conceptualizar como la razón a la que el objeto se desplaza a través del espacio-tiempo.

Introduciremos primero una nueva y útil notación que puede intimidar un poco al principio, pero que pronto asimilaremos. Se trata de incluir el tiempo en nuestras coordenadas espacio-temporales, transformando el índice latino i (que seguía el orden numérico 1, 2, 3) al índice griego μ (mu), que sigue el orden numérico 0, 1, 2, 3. Esto significa que asignamos al tiempo la «coordenada cero»: $t = x^0$. Podría pensarse que sería más natural reservarle al tiempo la cuarta coordenada (en lugar de la «ceroava»), pero designarlo como x^0 nos facilitaría la vida si alguna vez quisiéramos contemplar la posibilidad de un mayor (o menor) número de dimensiones

espaciales. El conjunto de coordenadas espacio-temporales queda por tanto así:

$$x^{\mu} = (t, x, y, z) = \left(x^0, x^1, x^2, x^3\right) \qquad (6.12)$$

La cuadrivelocidad de un objeto será la derivada de sus coordenadas en el espacio-tiempo.

Pero, un momento. Ha tenido que soportar ya una diatriba acerca del sinsentido que suponía hablar de la «velocidad del tiempo» en el mismo sentido en que hablamos de la velocidad a través del espacio. Entonces, ¿cómo podemos hablar ahora de velocidad a través del espacio-tiempo?

El quid está en definir la cuadrivelocidad como una derivada con respecto al tiempo propio τ a lo largo de la trayectoria, en lugar de hacerlo con respecto a la coordenada temporal t. Formulamos los componentes de la cuadrivelocidad así:

$$V^{\mu} = V^0, V^1, V^2, V^3 = \frac{dx^{\mu}}{d\tau} = \left(\frac{dt}{d\tau}, \frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau}\right) \qquad (6.13)$$

¿Qué representa físicamente la cuadrivelocidad? Por (6.8) sabemos que:

$$d\tau = \sqrt{1 - v^2}\, dt \qquad (6.14)$$

(Mostramos esto para trayectorias rectilíneas, pero es válido en general para infinitesimales, siendo v la velocidad en el punto a considerar). Así, los componentes de la cuadrivelocidad se convierten en:

$$V^{\mu} = \frac{1}{\sqrt{1-v^2}}\left(1, \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) \qquad (6.15)$$

Cuando v es muy pequeña, el factor es aproximadamente igual a 1 y puede ignorarse. Entonces el componente ceroavo V^0 de la cuadrivelocidad es simplemente 1, y los demás componentes son iguales a los de la trivelocidad corriente . En ocasiones lo expresaremos: (válido para velocidades pequeñas).

Ahora es cuando viene la magia de verdad. En la mecánica newtoniana prerrelativista, el momento se representa mediante un trivector, resultado del producto de la masa por la velocidad: . En el mundo de la relatividad, definimos el cuadrimomento de un modo análogo, como el producto de la masa por la cuadrivelocidad:

$$p^{\mu} = \left(p^0, p^x, p^y, p^z\right) = mV^{\mu} \qquad (6.16)$$

Al compararlos con (6.15), vemos que los componentes espaciales son muy similares a los de la expresión newtoniana para el momento, con un factora añadido de $1/\sqrt{1-v^2}$. ¿Y el componente temporal? Tenemos:

$$p^0 = \frac{m}{\sqrt{1-v^2}} \qquad (6.17)$$

De modo que el componente temporal del cuadrimomento es simplemente la masa del objeto dividida por un factor dependiente de la velocidad.

Como de costumbre, ganamos cierta intuición al considerar velocidades pequeñas (el «límite no relativista»). En este punto, recurramos a una poderosa estratagema matemática. Siempre que tenemos una expresión

(1+x)n, donde x es un número pequeño y n es algún exponente fijo, se obtiene una buena aproximación mediante:

$$(1+x)^n \approx 1+nx+\cdots \qquad (6.18)$$

Los puntos suspensivos indican que hay más términos (un número infinito, cuando n no es un entero positivo), pero serán proporcionales a x^2, x^3, y potencias más altas de x. Dado que hemos supuesto que el valor de x es muy pequeño, esas potencias más altas serán aún más pequeñas y pueden ignorarse.

Nuestro ubicuo factor de $1/\sqrt{1-v^2}$ es precisamente de esta forma, donde $x = -v^2$ y $n = -1/2$. (El valor de la raíz cuadrada de una cantidad es el mismo que el resultante de elevarla a la potencia ½, y la recíproca de una cantidad equivale a elevarla a la potencia 1). Así, para velocidades pequeñas tenemos:

$$\frac{1}{\sqrt{1-v^2}} \approx 1+\frac{1}{2}v^2 \qquad (6.19)$$

Conectando esto con (6.17), obtenemos:

$$p^0 \approx m+\frac{1}{2}mv^2 \qquad (6.20)$$

Hmm, esta última parte nos suena familiar. Es la energía cinética. Aparentemente, el componente ceroavo del cuadrimomento es algo de tipo energía que incluye un término constante m al que se suma la energía cinética.

Una idea: definamos en la relatividad la energía de un objeto simplemente como ese componente ceroavo del cuadrimomento:

$$E = p^0 = \frac{m}{\sqrt{1-v^2}} \qquad (6.21)$$

Como beneficio colateral, esta expresión nos ayuda a entender por qué es imposible acelerar un cohete hasta superar la velocidad de la luz. A medida que v se aproxima más y más a 1, el valor de $\sqrt{1-v^2}$ se aproxima a cero, y $E = m/\sqrt{1-v^2}$ tiende a infinito. Se necesitaría una cantidad infinita de energía para acelerar un objeto de masa finita hasta alcanzar la velocidad de la luz, no digamos ya superarla. Es algo que no va a suceder.

Cuando la velocidad es muy inferior a la de la luz, podemos servirnos de (6.20) para formular:

$$E \approx m + \frac{1}{2}mv^2 \qquad (6.22)$$

Podemos definir la energía cinética como «la energía del objeto debida al movimiento», y el otro término, que es simplemente la masa m, como «la energía del objeto cuando está incluso en reposo». Llamémosla energía en reposo: $E_{reposo} = m$.

Las unidades no están del todo bien, presumiblemente porque hemos partido de c = 1. Por la fórmula de la energía cinética sabemos que las unidades de la energía son las correspondientes al producto de la masa por el cuadrado de la velocidad. De modo que podemos fijar las unidades multiplicando por c^2. Lo cual nos lleva a la celebrada conclusión:

$$E_{\text{reposo}} = mc^2 \qquad (6.23)$$

Lo más frecuente es que vea esta fórmula sin la palabra «reposo» como subíndice, pero eso es porque nosotros estamos siendo muy cuidadosos y hay otras personas que a veces son más descuidadas. La forma correcta de entender esta fórmula es teniendo en cuenta que los objetos tienen energía aun cuando están completamente en reposo, y que la energía en reposo es igual al producto de la masa por el cuadrado de la velocidad. (O también puede pensar en la masa como en «el valor del cuadrimomento en el marco en reposo del objeto». Cualquiera de las dos formulaciones es válida).

Este es el ejemplo más famoso de la unificación conceptual proporcionada por la relatividad especial. La energía y el momento no son dos ideas distintas: la energía es la versión de tipo temporal del momento. Una de las características más maravillosas de la física es el modo en que unas nociones dispares pueden reunirse por el poder de una buena teoría.

1 El trabajo de Lorentz es tan fundamental para la relatividad especial, que a la teoría se la denominó ocasionalmente en un primer momento «teoría de Lorentz-Einstein». Pero Lorentz continuó aferrado a la idea de que existía un éter luminífero que definía un marco de reposo absoluto.

2 Los superíndices tienen un aspecto similar al de los exponentes, pero en este caso no es más que una etiqueta puesta sobre los tres componentes. Quizá se pregunte por qué no usamos subíndices, pero cuando pasemos a la relatividad general, necesitaremos usar ambos tipos, y representarán cosas ligeramente diferentes.

SIETE

GEOMETRÍA

Cuando Hermann Minkowski sugirió en 1907 que la mejor forma de concebir la relatividad especial era en términos de un espacio-tiempo unificado de cuatro dimensiones, Einstein se mostró escéptico. Se lamentó en una publicación de que la perspectiva de Minkowski «planteara grandes exigencias al lector en sus aspectos matemáticos».

Pero Einstein no tardó en apreciar el formalismo espacio-temporal, en especial al volver su atención hacia la cuestión de cómo integrar la gravitación en la relatividad. Finalmente, adquirió el convencimiento de que la podía entenderse la gravedad como una manifestación de la curvatura del propio espacio-tiempo. Es una visión que aporta una nueva percepción extraordinaria, pero no puede valer como teoría científica si no se traduce en un conjunto apropiado de ecuaciones. Tales ecuaciones iban a venir de la geometría, en particular de la geometría de Riemann, que permite que los espacios se curven arbitrariamente y se estudien desde dentro, en lugar de exigirles que encajen en un espacio de mayor dimensionalidad.

El problema era que Einstein no sabía nada acerca de la geometría de Riemann, al igual que casi ningún físico de la época. Era una cuestión tan solo iniciada en la década de 1850, y entrada la década de 1910 no había suscitado ninguna atención especial entre los físicos. Por fortuna, uno de los antiguos compañeros de clase de Einstein, Marcel Grossmann, era profesor de matemáticas y un experto en la materia. Bajo la tutela de Grossmann, Einstein aprendió suficiente geometría para formular la relatividad general, su teoría de la gravedad.

Ahora nos toca nosotros. Si Einstein tuvo que dejar a un lado su trabajo para estudiar la geometría de Riemann, ¿quiénes somos nosotros para resistirnos? El tema es lo suficientemente complejo y desconocido como

para dedicarle un capítulo entero —no caben demasiadas dudas acerca de que se trate de una Gran Idea—, y en el capítulo siguiente lo pondremos al servicio de la física.

LA GEOMETRÍA EUCLIDIANA

Todos tenemos recuerdos, más o menos felices, de la geometría que aprendimos en la secundaria. Había un montón de triángulos, círculos y demás figuras. El tipo de geometría que estudiamos entonces estaba asociada con Euclides, un matemático de la Antigüedad que estuvo activo en Alejandría no mucho tiempo después de que Aristóteles enseñara en Atenas. Podríamos calificarla de geometría «de mesa», por cuanto habla de propiedades de líneas y curvas que dibujamos sobre una superficie plana, de dos dimensiones. (Aunque es fácil generalizar sus principios a tres o más dimensiones espaciales).

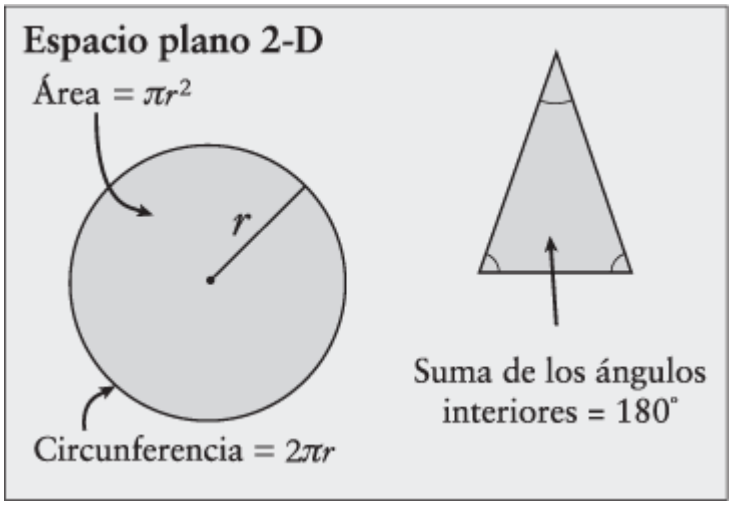
Lo que hizo de Euclides tan influyente no es tanto ningún resultado en particular —teoremas acerca de las propiedades de los objetos geométricos —, cuanto el enfoque que instauró. En realidad, la geometría euclidiana abarca buen número de resultados clásicos, muchos de los cuales son anteriores al trabajo de Euclides:

•Teorema de Pitágoras: el cuadrado de la longitud de la hipotenusa de un triángulo rectángulo es igual a la suma de los cuadrados de los otros dos lados del triángulo.

•Los ángulos interiores de un triángulo suman 180 grados (π radianes).

•La circunferencia de un círculo es $2\pi r$, donde r es el radio del círculo.

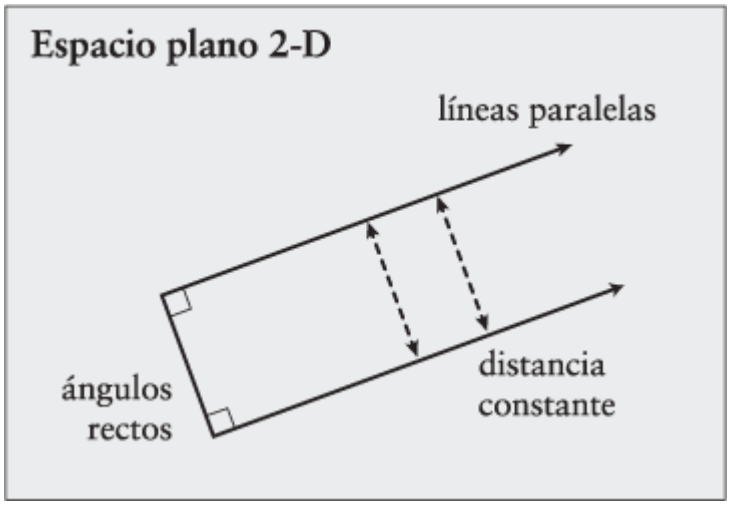
•El área del interior de un círculo (el «disco») es π r 2.



El rasgo distintivo de la geometría euclidiana es el de definir un sistema axiomático: establecemos un conjunto de postulados, o axiomas, y a partir de ellos utilizamos las reglas de la lógica para obtener teoremas. Con tal de que lo hagamos apoyándonos en las reglas básicas de la lógica —como tiene que ser, aunque por supuesto hay sutilezas que siempre preocupan a los filósofos—, los teoremas serán tan verdaderos como los axiomas. Uno no tiene necesariamente que tomar partido acerca de si los axiomas son verdaderos o no: un teorema es un enunciado que dice que «si estos axiomas son verdad, este resultado también lo es».

Se trata de un proceso de elaboración muy diferente del método científico. La ciencia es empírica y falibilista: cualquiera de nuestras teorías científicas podría ser errónea, por numerosas que sean las pruebas acumuladas hasta el momento a favor de ella. Los científicos sugieren hipótesis acerca de cómo funciona el mundo, las contrastan con datos y les conceden el crédito debido en consecuencia. Uno nunca puede estar absolutamente seguro de que su hipótesis sea correcta, pues mañana podrían llegar nuevos datos que nos hicieran cambiar de opinión. Pero en geometría, y en general en matemáticas y en lógica, uno puede estar seguro de que si sus axiomas son verdaderos, sus teoremas también lo serán.[1]

En su mayor parte, los axiomas de Euclides son enunciados razonables, cuyo servicio para la fundamentación para la geometría tiene todo el sentido. Ejemplo de ello sería: «podemos trazar una línea recta entre dos puntos cualesquiera», o «todos los ángulos rectos son iguales entre sí». Pero hay uno que siempre destacó como diferente: el quinto postulado de Euclides, conocido como postulado de las paralelas. Una manera de formularlo consistiría en trazar en un plano dos líneas inicialmente paralelas, por ejemplo, partiendo de un segmento y dibujando dos líneas que salen de él en sendos ángulos de 90 grados. De acuerdo con el postulado de las paralelas, esas líneas inicialmente paralelas mantendrán siempre una separación equidistante entre sí. (La forma en que Euclides planteó el enunciado vendría a ser: «si los ángulos son más pequeños que el ángulo recto, las líneas terminarán por encontrarse»).



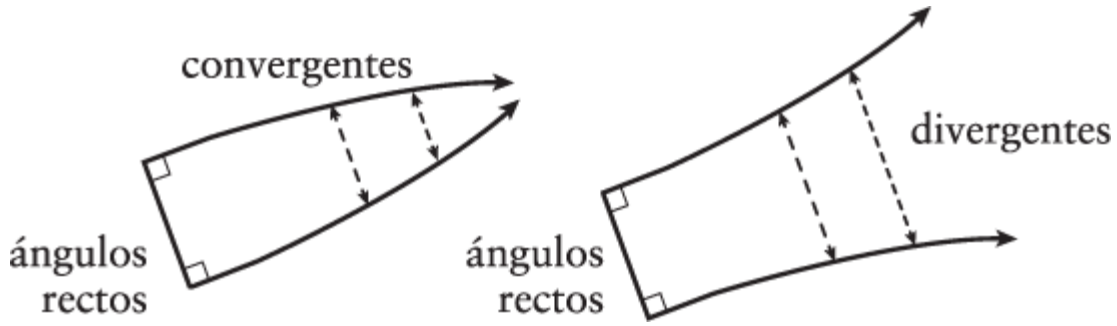
Esto también parece bastante razonable, de acuerdo con nuestra intuición acerca de cómo funcionan las cosas en un plano. (Los matemáticos imaginan planos que se extienden infinitamente en todas direcciones, pero a nosotros nos basta pensar de forma intuitiva en la superficie de una mesa o una hoja de papel). Reconoceremos, no obstante, que suena un poco más burdo que los demás postulados. Durante muchos años, los geómetras pensaron que sería posible demostrar el postulado de las paralelas a partir de los demás axiomas de Euclides, lo cual resultaría en desplazarlo del estante de «axiomas» al de «teoremas». Cuando cursé geometría en el

instituto, nuestro profesor nos troleó con la promesa de subirle la nota a quien lo demostrara correctamente. Nadie lo consiguió.

GEOMETRÍA NO EUCLIDIANA

No le gastaré la misma jugada. No le insinuaré que demuestre el postulado de las paralelas a partir del resto de axiomas de la geometría euclidiana, porque no es posible hacerlo. Lo sabemos porque podríamos sustituir el postulado de las paralelas por otro postulado diferente y, al añadirle a este los demás axiomas, obtendríamos versiones distintas de la geometría, todas ellas perfectamente coherentes por derecho propio.

Es bastante fácil intuir qué forma podrían adoptar los postulados alternativos. Si el postulado de las paralelas dice que dos líneas inicialmente paralelas mantendrán la misma distancia entre ellas por siempre, las alternativas dirán presumiblemente que no mantendrán la misma distancia. Hay dos formas en que tal cosa podría pasar: que las líneas se junten o que se separen.



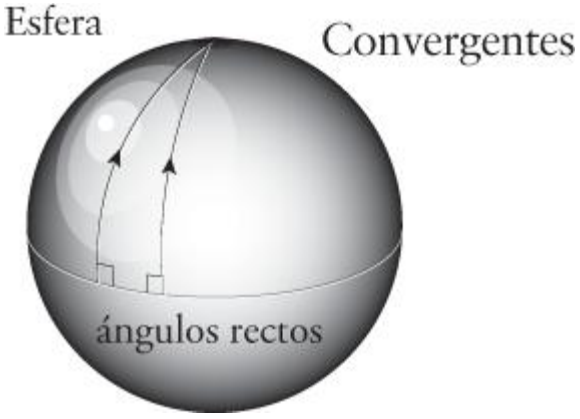
Si le preocupa que las líneas rectas reales no se comportan así, no tema. Nuestra intuición euclidiana se debe a estar familiarizados con la geometría sobre un plano o equivalente. Por algo los planos son planos, no curvos. Se supone que estos postulados alternativos son aplicables a algún tipo de espacio bidimensional, pero un espacio provisto de curvatura, en lugar de

un plano. Nuestros postulados alternativos definen algún tipo de geometría diferente, oportunamente llamada geometría no euclidiana.

Esto no significa que los sistemas geométricos basados en estos postulados sean enteramente abstractos e hipotéticos. Al fin y al cabo, estamos refiriéndonos a formas diferentes a las de un plano, pero bidimensionales como, por ejemplo, una esfera. (Al hablar de una esfera nos referimos exclusivamente a su superficie, sin el interior). Podríamos preguntarnos qué sucedería con dos líneas inicialmente paralelas en una esfera.

Quizá se pregunte qué entendemos exactamente por «línea», ya que nada que tracemos sobre la superficie de una esfera va a parecer perfectamente recto. Por el momento, no piense más que en círculos mayores o en porciones de estos. Un círculo mayor sería la curva formada por la intersección de la esfera con un plano que pasa por su centro. El ecuador es un ejemplo de ello, al igual que las líneas de longitud (¡no de latitud!), pero podemos también imaginar círculos mayores inclinados en un ángulo cualquiera.

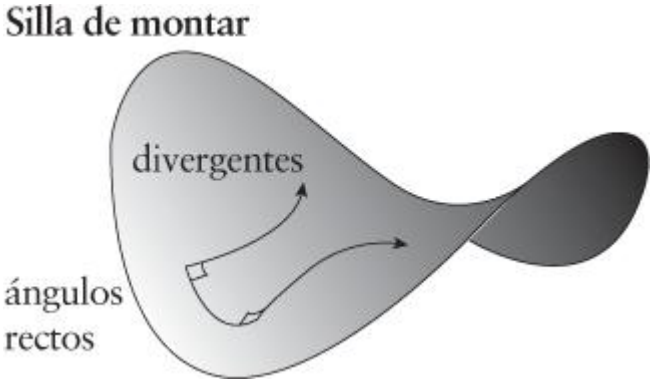
Consideremos, pues, un segmento del ecuador de una esfera y dos líneas que parten hacia el norte en ángulo recto. Que sean todo lo rectas posible, es decir, que formen círculos mayores. Nos guste o no, se van a encontrar (en este caso, en el polo norte).



Hay otras estimadas características de la geometría euclidiana que cojean cuando pasamos de un plano a una esfera. Piense en círculo de radio r, con

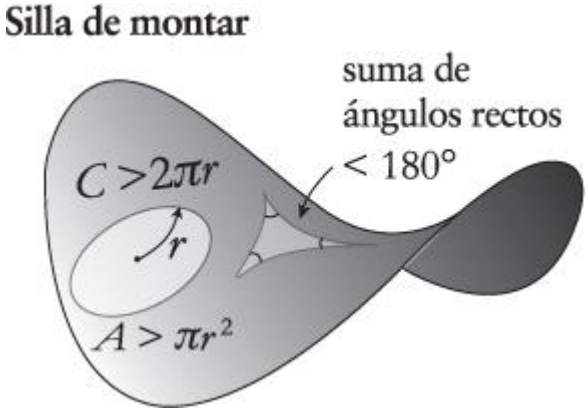
centro en el polo norte. Podemos apreciar en la figura cómo la circunferencia de este círculo es menor a $2\pi r$, y el área del círculo es menor a πr^2. (Si prolongamos el radio hasta el polo sur, la circunferencia sería cero). Por otra parte, los ángulos interiores de un triángulo sumarán por lo general más de 180 grados. De hecho, podemos construir un triángulo con tres ángulos interiores de 90 grados en la esfera, uniendo tres segmentos que midan una cuarta parte de sendos círculos mayores. Euclides debe de estar removiéndose en su tumba.

¿Y en el caso de líneas inicialmente paralelas que luego se separen, en lugar de acercarse? Esto es más difícil de visualizar, pero sería comparable a la figura descrita por una silla de montar, o por un chip de patatas Pringles.



En una geometría de formas similares a sillas de montar, las propiedades de los círculos y los triángulos resultan modificadas también, pero en la

dirección opuesta al caso de la geometría de esferas. Dado un círculo de radio r, su circunferencia será mayor que $2\pi r$, el área del círculo mismo será mayor que πr^2, y los ángulos interiores de un triángulo sumarán generalmente menos de 180 grados.



En cada uno de estos casos subyace implícito un supuesto simplificador: el de que cualesquiera que sean las propiedades geométricas de nuestro espacio bidimensional, siguen siendo las mismas en todo lugar y en todas direcciones. Si tomamos un segmento de línea de una longitud determinada y trazamos líneas inicialmente paralelas perpendiculares a él, las líneas se separarán o convergerán en la misma proporción, sea cual sea el lugar en que situemos el segmento o la dirección en la que apunte. La expresión técnica que utilizamos en este caso es que estamos considerando geometrías de curvatura constante, por oposición a curvaturas que cambian en función de la ubicación o de la dirección. Esto facilita muchos las cosas, lo cual significa que pronto vamos a renunciar a este supuesto. Pero proporciona un punto de partida natural para un paseo que no tardará en convertirse en una aventura bastante alocada.

Habiendo asumido el supuesto de la curvatura constante, solo hay estas tres opciones para la geometría en un espacio de dos dimensiones: que las líneas inicialmente paralelas se mantengan paralelas, que sean convergentes, o que sean divergentes.

•Si se mantienen paralelas: geometría euclidiana. Un plano (curvatura cero).

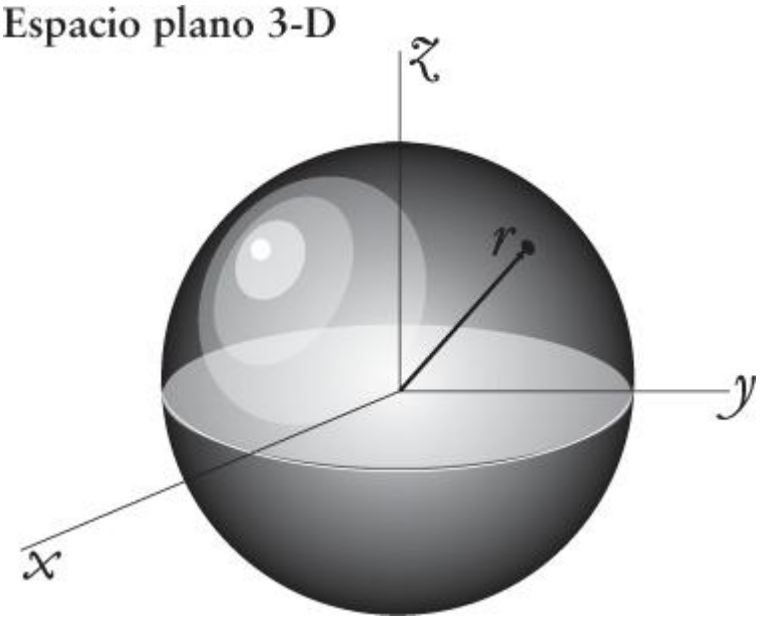
•Si son convergentes: geometría esférica (o elíptica). Curvatura positiva.

•Si son divergentes: geometría hiperbólica. Curvatura negativa.

INTRÍNSECO Y EXTRÍNSECO

La idea de una geometría no euclidiana no entró en escena hasta comienzos del siglo XIX, más de dos mil años después de la geometría euclidiana. Nikolai Ivanovich Lobachevsky en Rusia y János Bolyai en Hungría desarrollaron independientemente las ideas básicas de la geometría hiperbólica. Puede parecer sorprendente, tanto que la geometría hiperbólica se conceptualizara antes que la esférica, como que, para empezar, tardara tanto tiempo en aparecer. ¿Tan difícil es imaginar figuras geométricas trazadas en una esfera?

Ambas preguntas —por qué tardó tanto tiempo en surgir y por qué los matemáticos dieron primero con la geometría hiperbólica— se contestan en parte una a la otra. Por supuesto que todo el mundo conocía las esferas y se hacía una idea del comportamiento de líneas, círculos y ángulos sobre su superficie. Pero no pensaban en las esferas como en algo que definiera un tipo de geometría por separado. La superficie esférica bidimensional se consideraba inserida en el espacio euclidiano tridimensional. De hecho, una forma de definir la esfera es como «el conjunto de todos los puntos que se encuentran a cierta distancia R de un punto determinado en el espacio». Nadie sentía la motivación de crear toda una clase nueva de geometría para describir las esferas. Sus propiedades podían derivarse de la vieja y estupenda geometría euclidiana en tres dimensiones, aplicada en particular a esta forma inserida en ella.



No es el mismo caso para superficies curvadas en sentido negativo. Podemos concebir sillas de montar o chips de patata, por supuesto, pero un examen detenido revela que, mientras que estos ejemplos del mundo real presentan una curvatura negativa, tal curvatura no es constante en todas partes. Hay un punto central bien definido, a partir del cual, y a medida que nos alejamos de él, la curvatura declina gradualmente. No depende de lo inteligentes que seamos, es imposible insertar fielmente un espacio bidimensional de curvatura negativa constante en un espacio euclidiano tridimensional.

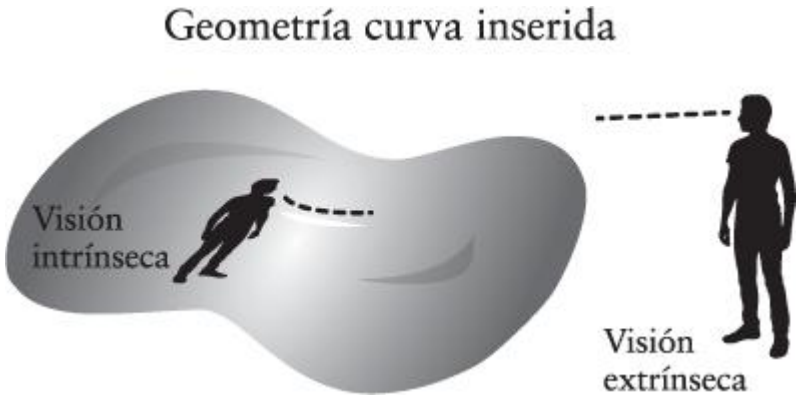
Esta es la razón de que la invención de la geometría hiperbólica constituyera un logro intelectual verdaderamente impresionante. Podemos postular un conjunto de axiomas que integren las propiedades geométricas del espacio bidimensional de curvatura negativa constante, a veces llamado plano hiperbólico (en contraposición al plano «plano» de Euclides). Podemos demostrar teoremas basados en este sistema, derivar fórmulas para el perímetro de una circunferencia y el área del círculo que encierra, y responder a todas las demás preguntas geométricas que puedan plantearse. Lo que no podemos es construir un espacio así como una superficie bidimensional inserta en el espacio en que vivimos. Un plano hiperbólico exacto solo existe en nuestras mentes.

Esto supone un salto conceptual muy grande, por lo que respecta a la historia de las matemáticas. Libres de la limitación autoimpuesta de considerar únicamente objetos que podamos construir en la realidad, los matemáticos podían estudiar las implicaciones de todo tipo de sistemas axiomáticos por su interés como tales.

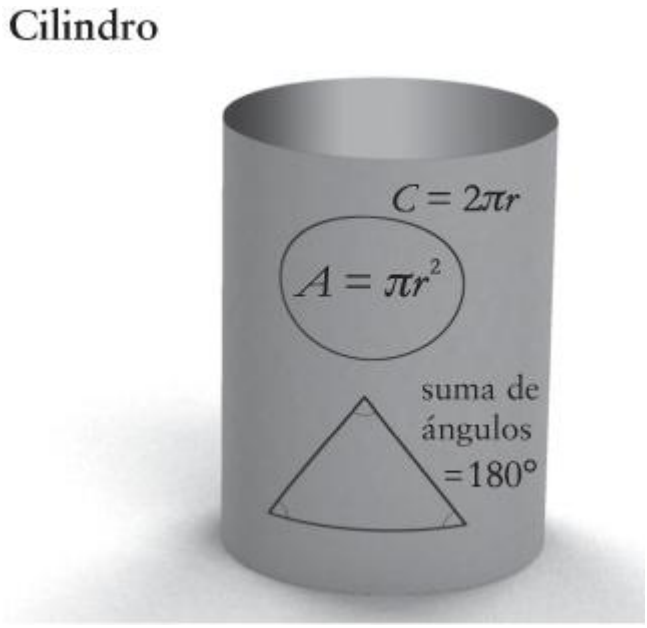
Para el propósito que nos ocupa, la idea principal que se desprende de ello es que podemos estudiar las propiedades geométricas intrínsecas de un espacio, en oposición a las propiedades extrínsecas que hereda por estar inserto en un espacio mayor. Esta distinción fue desarrollada por Carl Friedrich Gauss, uno de los más grandes matemáticos de todos los tiempos. (Gauss, famoso por lo que tardaba en poner por escrito sus resultados, reclamaba para sí el mérito de haber ideado la geometría hiperbólica antes que Lobachevsky y Bolyai, pero nunca publicó nada sobre ello, y a uno no le reconocen las buenas ideas que se guarda para sí mismo).

Cuando observamos formas dibujadas sobre la superficie de una mesa, o una esfera en un espacio tridimensional, las percibimos naturalmente desde el exterior. La «curvatura», desde esta perspectiva, nos muestra cómo se dobla y retuerce tal forma dentro del espacio mayor en que está inserida. Esto es la curvatura extrínseca.

Pero podemos también imaginar cómo serían las cosas percibidas por seres imaginarios que vivieran en el interior de dicha forma. Podrían dibujar un círculo y medir su circunferencia, por ejemplo. El resultado no tiene nada que ver con que la forma esté inserida en un espacio mayor; ni siquiera se necesita que exista tal tipo de espacio. Las características que pueden medirse puramente desde el interior definen la geometría intrínseca del espacio.



Como ejemplo elocuente de la distinción entre curvatura intrínseca y extrínseca, podemos considerar un cilindro bidimensional inserto en un espacio tridimensional.



Ciertamente, un cilindro parece curvo, pero eso es una construcción de nuestra perspectiva externa. Lo que percibimos nosotros es la curvatura extrínseca, pero intrínsecamente el cilindro es perfectamente plano. Podemos comprobarlo si trazamos imaginariamente dos líneas rectas inicialmente paralelas y observamos si convergen o se separan, o si dibujamos un triángulo y sumamos sus ángulos interiores, o verificando la

relación entre el radio de un círculo y su circunferencia o su área. En cualquiera de estos casos, la geometría intrínseca del cilindro es la misma que la de un plano extendido.

Estamos demorándonos en esta cuestión porque terminaremos hablando de la curvatura del espacio-tiempo mismo. Nuestro universo no está inserto en un espacio mayor; al menos no tiene por qué estarlo, hasta donde sabemos. En lo que al espacio-tiempo se refiere, la geometría intrínseca es todo lo que tenemos.

VARIEDADES

Además de percibir la distinción entre curvatura intrínseca y extrínseca, Gauss fue también pionero en el estudio de casos en que la curvatura no es constante: espacios en que la curvatura varía en cantidad y forma de un lugar a otro. Pero cuando llegó el momento de desarrollar en todo su alcance la teoría de tales geometrías arbitrarias, dejó la tarea en manos de su discípulo Bernhard Riemann.

En 1853, Riemann estaba en la etapa final de sus estudios, a falta de la impartición de una clase magistral para la obtención de su habilitación, grado del sistema universitario alemán, posterior al doctorado, requisito habitual para poder impartir clases en la universidad. (En su trabajo doctoral, Riemann había explorado el uso de números complejos para el estudio de superficies bidimensionales, y el grueso de las investigaciones conducentes a su habilitación lo había dedicado a la formulación más rigurosa del cálculo integral. Era un joven ambicioso y productivo).

Continuando con la historia, Riemann propuso una lista de temas posibles para su clase magistral a Gauss, quien sorprendió a su alumno escogiendo el tema que a Riemann le parecía el menos interesante de todos: los fundamentos de la geometría. El cometido obligó a Riemann a volver a sentarse en su estudio y reflexionar con intensidad acerca de lo que realmente queremos decir cuando hablamos de «la geometría» de un

espacio, y de hecho de qué tipo de «espacio» estamos hablando, especialmente desde el punto de vista del observador intrínseco. (Al comienzo de la clase se lamentó de «no ser ducho en tareas de naturaleza filosófica como aquella», pero se las arregló más que bien). Terminó escribiendo uno de los artículos más influyentes de la historia de las matemáticas, cuyos resultados ocupan todavía un lugar central en la relatividad general y la visión moderna del espacio-tiempo. Por desgracia, Riemann murió aún joven, de tuberculosis, de lo contrario podría haber extraído un mayor número de resultados fundamentales para las matemáticas.

Riemann comenzaba su disertación definiendo la noción de variedad: un conjunto infinito de puntos conectados diferencialmente con un espacio de cierta dimensionalidad definida. Recuerde que cuando realizamos una ampliación de la imagen en primer plano, una curva parece una línea recta. Las variedades se construyen partiendo de esta misma noción, pero en más de una dimensión. En un espacio curvo, si la ampliación es lo suficientemente cercana, las cosas parecen responder a la geometría euclidiana. La curvatura que se manifiesta en escalas mayores, puede codificarse en la forma en que los bits infinitesimales de espacio plano se unen globalmente.

Lo que tiene una importancia crucial en lo que respecta a las variedades, es que no necesitan inserirse en ningún otro espacio, aunque a veces las dibujemos de manera que parece que lo estén. Cuando representamos variedades de dos dimensiones como una esfera o un toro, estamos pensando en la superficie bidimensional, no en el espacio tridimensional en que están inseridas. Esto es producto simplemente del modo en que nosotros, criaturas tridimensionales, representamos las cosas. Las

variedades tienen una topología y una geometría bien definidas por derecho propio. Hay que pensar siempre intrínsecamente, no extrínsecamente.

GENERALIZACIÓN DEL TEOREMA DE PITÁGORAS

A continuación, tendríamos que decidir cómo especificar la geometría de una variedad, refiriéndonos únicamente una vez más a nociones intrínsecas que pudieran ser medidas por parte de seres humanos que vivieran en su interior. Potencialmente hay muchas maneras de hacerlo. El objetivo es estipular algunas cantidades geométricas básicas de la variedad, a partir de las cuales pueda derivarse cualquier otra que queramos calcular.

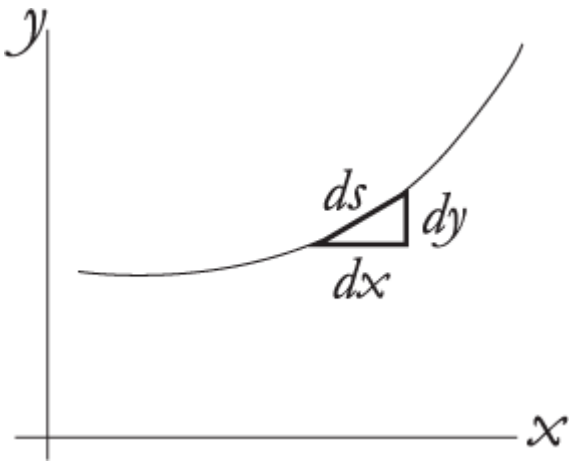
Las cantidades establecidas por Riemann fueron las longitudes de las curvas. No de una sola curva en especial, sino de cualquier curva que quisiéramos dibujar en el interior de la variedad. Si conocemos la longitud de cualquier curva posible que podamos dibujar, a partir de ahí podemos obtener cualquier cosa que queramos saber acerca de la geometría.

Parece una labor intimidatoria, especificar la longitud de todas las curvas posibles. Incluso en variedades simples como la esfera o el plano, hay un montón abrumador de curvas que podemos dibujar. Por suerte contamos con una herramienta para manejarnos con fieras tan indómitas: el cálculo. No necesitamos anotar literalmente la longitud de cada curva, solo tenemos que establecer una fórmula para calcular la longitud de una porción infinitesimal de una curva. Luego podemos obtener la longitud entera integrando dicha porción infinitesimal.

Para captarlo intuitivamente, consideremos un plano bidimensional con un sistema cartesiano de coordenadas (x, y). Alguien dibuja una curva sobre el plano, y nosotros queremos calcular la longitud ds de una porción infinitesimal de la curva, definida por desplazamientos infinitesimales en las coordenadas, dx y dy. Ya conocemos la fórmula para ello: no es más que el teorema de Pitágoras. [Si la siguiente formulación le genera imágenes retroactivas de la ecuación (6.10) para la longitud de un segmento de tipo

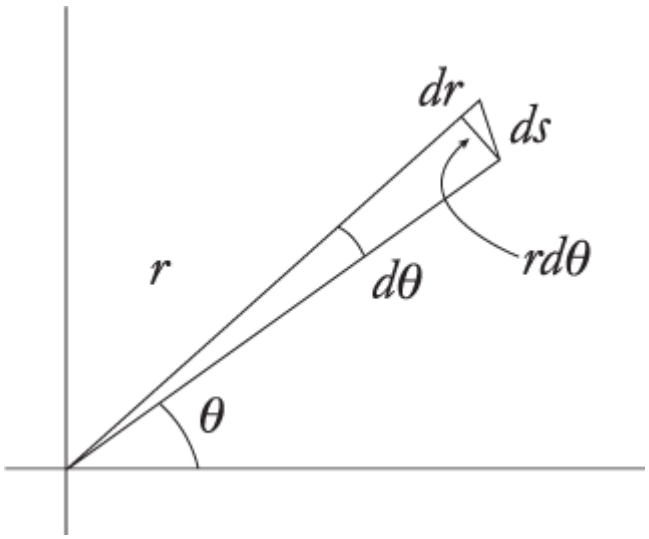
espacial en el espacio-tiempo de Minkowski, está hecho con toda intención].

$$ds^2 = dx^2 + dy^2 \qquad (7.1)$$



Está bastante claro. Pero las coordenadas cartesianas son muy especiales. Las líneas de una x constante son todas perfectamente rectas y siempre paralelas entre sí, al igual que las líneas de una y constante. Si el postulado de las paralelas no se sustenta —si nuestra variedad es no euclidiana—, no podemos dibujar coordenadas como estas en todas partes. No hay coordenadas cartesianas que valgan para cubrir una esfera, por ejemplo.

Aun cuando nuestra variedad es plana, no estamos obligados a utilizar coordenadas cartesianas. Piense en unas coordenadas polares en un plano, donde hemos especificado un punto con una distancia r con respecto al origen de coordenadas y un ángulo θ con respecto al eje horizontal. Considere la longitud de un segmento infinitesimal en función de estas coordenadas.



La longitud física correspondiente a un cambo angular dθ no es una cantidad fija; crece más a medida que r se incrementa. La longitud de un segmento diminuto a la distancia r fijada no será dθ, sino rdθ. Está bastante claro cómo adaptarlo. La fórmula correcta para un segmento infinitesimal arbitrario es:

$$ds^2 = dr^2 + r^2 d\theta^2 \qquad (7.2)$$

Es casi el teorema de Pitágoras, pero no del todo. Hay un factor extra, r^2, delante del término $d\theta^2$. Este factor extra entraña la idea de que, a medida que r aumenta, se corresponde un incremento dθ de longitud física cada vez mayor. En esencia, conservamos el espíritu de Pitágoras, pero adoptando una fórmula ligeramente más general, que relacione distancias físicas con incrementos en los valores de las coordenadas.

EL ELEMENTO DE LÍNEA

Sirva este ejemplo de inspiración para tratar de entender una idea llamada elemento de línea, fórmula general por la que podemos tomar un segmento infinitesimal, definir desplazamientos infinitesimales en cada una de las coordenadas y utilizar estos desplazamientos para calcular la longitud ds del segmento. Partiremos de variedades bidimensionales para mayor simplicidad.

Imaginemos que tenemos dos coordenadas (x^1, x^2), entendiendo los números volados como superíndices, no como exponentes. (Tal como hicimos con los componentes de vectores). Queremos asociar la distancia ds a los desplazamientos de coordenada dxi. Siguiendo nuestra inspiración pitagórica, esperamos poder asociar el cuadrado de la distancia ds^2 a los cuadrados de los desplazamientos de coordenada, como $(dx^1)^2$ —aquí el 2 indica el cuadrado de dx^1, nos vemos obligados a utilizar los números volados como superíndices y como exponentes en una misma expresión—. Y para establecer una adecuada generalización, deberíamos introducir «términos cruzados» que multipliquen coordenadas separadas, como dx^1dx^2. Finalmente, tal como aprendimos de las coordenadas polares, puede haber coeficientes que multipliquen estos términos, los cuales pueden depender a su vez de las coordenadas.

La fórmula de este tipo más general para los dos elementos de línea de dos dimensiones adopta la forma:

$$ds^2 = A\left(x^1, x^2\right)\left(dx^1\right)^2 + B\left(x^1, x^2\right)dx^1dx^2 + C\left(x^1, x^2\right)\left(dx^2\right)^2 \tag{7.3}$$

Son muchos paréntesis, exponentes y superíndices, de modo que tomemos un respiro y observemos por dónde vamos. Las tres cantidades A, B y C son números que asumen valores específicos en cada punto de la variedad, por lo que los hemos expresado como funciones de las coordenadas (x^1, x^2). Cada uno de ellos multiplica un producto de dos de los desplazamientos de coordenada, dx^1 y dx^2. Tenemos así no solo un término proporcional a $(dx^1)^2$

y otro proporcional a $(dx^2)^2$, sino también un nuevo término proporcional a dx^1dx^2. Esto será relevante si las coordenadas no son perpendiculares.

Por esto es tan importante esta expresión: una vez dadas las tres funciones $A(x^1, x^2)$, $B(x^1, x^2)$ y $C(x^1, x^2)$, la fórmula (7.3) nos proporciona un modo de calcular la longitud de cualquier curva que dibujemos. Y esta información, dice Riemann, es suficiente para determinar completamente la geometría de la variedad. Todo cuanto queramos llegar a saber —ángulos, áreas, curvatura— está contenido en estas tres únicas funciones. Algo similar puede decirse para más de dos dimensiones, aunque el número de funciones sería un poco más largo. En un espacio de d dimensiones, se necesitan d(d + 1)/2 funciones para especificar por entero la línea de elemento.

No es que vaya a ser sencillo extraer la información geométrica que nos interese a partir de estas funciones. El problema es que la línea de elemento puede adoptar un aspecto muy diferente para sistemas de coordenadas diferentes, aunque la geometría subyacente sea exactamente la misma. Ya lo hemos visto en el caso del plano simple. Para coordenadas cartesianas, el elemento de línea (7.1) expresado según la forma (7.3) sería:

$$A(x, y)=1, B(x, y)=0, C(x, y)=1 \qquad (7.4)$$

Mientras que para coordenadas polares, el elemento de línea (7.2) corresponde a:

$$A(r, \theta)=1,\ B(r, \theta)=0,\ C(r, \theta)=r^2 \qquad (7.5)$$

La misma geometría, pero con diferentes versiones del elemento de línea, porque estamos en sistemas de coordenadas diferentes. La geometría no se preocupa por sistemas de coordenadas; estas son inventos humanos, no características intrínsecas de la variedad. Vamos a tener que trabajar con un poco más de esfuerzo para exprimir la información acerca de la curvatura, a partir de estos elementos de línea.

LA MÉTRICA

Vamos a tener que introducir una vez más una nueva notación ad hoc. Las buenas notaciones son una recompensa en sí mismas, pero el número de funciones independientes necesarias para especificar el elemento de línea va a resultar muy pronto inmanejable para dimensiones de número más elevado.

Tampoco debería ser demasiado difícil, porque ya hemos visto el patrón a seguir en (7.3): por cada par de desplazamientos de coordenada dxi y dxj (donde i y j son superíndices del número correlativo de dimensiones, y pueden ser de valor igual o diferente), asignamos una función de espacio-tiempo. Podemos designar estas funciones como gij(x), donde la austera variable x representa ahora todas las coordenadas a la vez. Las letras i y j no tienen ningún significado en o por sí mismas; son simplemente etiquetas que denotan el valor de los índices, podemos utilizar las letras que queramos. Así, g11 es la función que multiplica a $(dx^1)^2$ en el elemento de línea, g12 es la función que multiplica a dx^1dx^2, y así sucesivamente.

Estas funciones pueden expresarse en forma de matriz,[2] por ejemplo, en tres dimensiones:

$$g_{ij} = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \qquad (7.6)$$

Este es el famoso tensor métrico, que será centro de nuestra atención cuando pasemos a la relatividad general. Cada entrada en la matriz, conocida como elemento de la métrica, es una función de posición por separado. Y sabemos que todos estos elementos nos dicen todo lo que

queremos conocer acerca de la geometría de la variedad a considerar. Lo más habitual es que, cuando un físico está tratando de resolver un problema de investigación en el ámbito de la relatividad general, o está tratando de calcular cuál es la métrica (por ejemplo, ante determinada distribución de materia y energía), o está indagando las consecuencias físicas de alguna forma particular de la métrica. (En la relatividad, nuestra métrica será sobre el espacio-tiempo, no solo sobre el espacio, pero la diferencia en la formulación matemática va a ser sorprendentemente pequeña. Simplemente, nos serviremos de letras griegas para designar los elementos).

Conocer la métrica es justamente equivalente a conocer el elemento de línea. La relación es simple:

$$ds^2 = \sum_{i,j} g_{ij} dx^i dx^j = g_{11}\left(dx^1\right)^2 + g_{12} dx^1 dx^2 + g_{13} dx^1 dx^3 + \cdots \tag{7.7}$$

De modo que si conocemos todos los elementos de la métrica, podemos calcular la longitud de cualquier curva. A partir de ahí, aunque es posible que no sea obvio, podemos calcular también áreas y volúmenes, ángulos entre líneas, y muchas otras cosas más.

Nuestro elemento de línea bidimensional (7.3) encaja en este patrón, como debe ser. En este caso tenemos:

$$g_{ij} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} A & \frac{1}{2}B \\ \frac{1}{2}B & C \end{pmatrix} \tag{7.8}$$

Vemos que B aparece dos veces, pero en ambas dividido por la mitad. Esto es porque una aplicación literal de la fórmula (7.7) nos da contribuciones diferentes para $dx^1\, dx^2$ y $dx^2\, dx^1$, pero aquellas son iguales entre sí. Se

requiere, por tanto, que la métrica sea simétrica: para cada i y j tenemos gij = gji. Desde el punto de vista de la versión matricial, los elementos de la parte superior derecha serán iguales a sus correspondientes en la parte inferior izquierda.

Pongamos los pies en el suelo con algunos ejemplos. Podemos comenzar por un terreno más familiar: el espacio tridimensional euclidiano con coordenadas cartesianas (x, y, z). Aquí sabemos cuál es el elemento de línea; proviene del teorema de Pitágoras.

$$ds^2 = dx^2 + dy^2 + dz^2 \qquad (7.9)$$

El tensor métrico, expresado en forma de matriz, no podría ser más simple:

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (7.10)$$

(En lo sucesivo, asumimos que el elemento superior izquierdo es g11, el que está inmediatamente a la derecha de este es g12, y así sucesivamente, así que no escribiremos esta notación explícitamente). Esta es la métrica euclidiana, que puede hacerse extensiva a cualquier otro número de dimensiones añadiendo más filas y columnas, manteniendo los 1 en la diagonal y los 0 en todas las demás posiciones. El propio Euclides la utilizaba implícitamente en todo momento, aunque no pensara en tales términos. La métrica define la geometría.

Pero existen también sistemas de coordenadas no cartesianos. Para coordenadas polares en dos dimensiones (r,θ), podemos extraer la respuesta de (7.2):

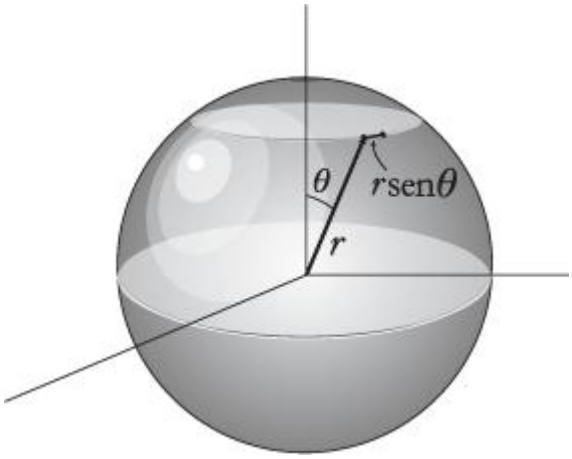
$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \qquad (7.11)$$

Debería estar claro que los elementos de una métrica son siempre relativos a determinado sistema de coordenadas. Una expresión como (7.11) solo tiene sentido si sabe que estamos utilizando coordenadas (r,θ), con $x^1 = r$ y $x^2 = \theta$.

Está también la versión tridimensional de las coordenadas polares, las coordenadas esféricas (r,θ,ϕ) que mencionamos en el capítulo 4. La métrica para un espacio plano en estas coordenadas es:

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \operatorname{sen}^2\theta \end{pmatrix} \qquad (7.12)$$

A medida que el ángulo va desde cero en el polo norte hasta 90 grados en el ecuador y luego hasta 180 grados en el polo sur, sen va de 0 a 1 y luego de vuelta a 0. Su aparición en el elemento g33, en la esquina inferior derecha de (7.12), refleja el hecho de que la distancia física asociada con un cambio en $x^3 = \phi$ es pequeña cerca de los polos y más grande cerca del ecuador.



Nunca insistiremos en demasía en que la geometría definida por (7.12) es la misma geometría que la definida por (7.10). Ambas son «métrica euclidiana» —espacio plano—, pero una está expresada en coordenadas cartesianas y la otra en coordenadas esféricas.

LA MÉTRICA EN ESPACIOS CURVOS

Observemos ahora una variedad con alguna curvatura. El ejemplo más sencillo es el de una esfera bidimensional. Por fortuna, es bastante fácil hallar la métrica en una esfera una vez hemos expresado la métrica (7.12) para un espacio plano en coordenadas esféricas. Una esfera es simplemente el subconjunto de espacio plano resultado de asignar un valor específico R a la coordenada radial r. Así, podemos obtener la métrica en una esfera a partir de (7.12), borrando la fila superior y la columna de la izquierda (ya que ambas se refieren a distancias en la dirección r, que no existe cuando estamos reducidos a una esfera), y estableciendo r = R. El resultado es la métrica sobre una esfera en coordenadas $(x^1, x^2) = (\theta,\phi)$:

$$g_{ij} = \begin{pmatrix} R^2 & 0 \\ 0 & R^2 \operatorname{sen}^2 \theta \end{pmatrix} \qquad (7.13)$$

Esto muestra un parecido familiar con (7.12), la métrica euclidiana tridimensional sobre coordenadas polares, pero son especies muy diferentes. Aquí R no es una coordenada, sino un parámetro fijo que nos dice cómo de grande es nuestra esfera. Y lo que es más importante, esta métrica no es plana. Supone un cierto reto observar simplemente una métrica y concluir si es plana o curva. Según la elección de coordenadas que se haga, puede oscurecerse la percepción de la geometría subyacente.

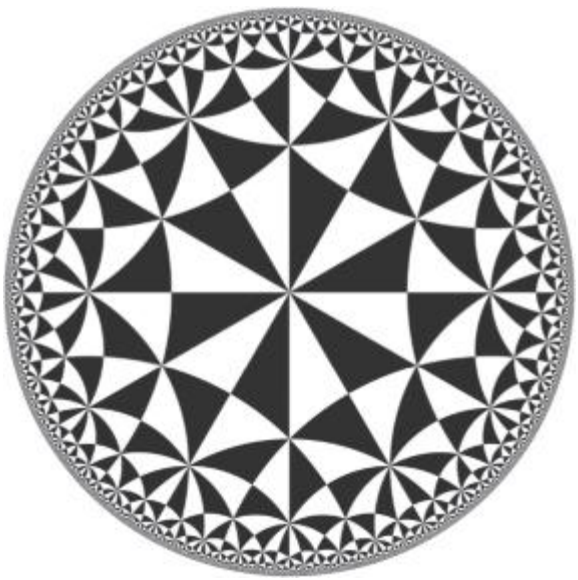
Volvamos a la métrica del plano hiperbólico bidimensional, siguiendo las investigaciones de Lobachevsky y Bolyai (aunque no lo llamaron así, no más de lo que lo hubiera hecho Euclides). Como siempre, la primera cuestión es la de decidir qué sistema de coordenadas utilizar. Un sistema bastante conveniente es el conocido como disco de Poincaré. (Eugenio Beltrami fue el primero en utilizarlo, Henri Poincaré lo utilizó más tarde, pero Poincaré es más famoso, y tendemos a nombrar las cosas asociándolas con los nombres de las personas famosas).

Generalmente se designan las coordenadas como $(x^1, x^2) = (x, y)$, lo cual nos hace pensar en coordenadas cartesianas sobre un plano, pero hay una restricción adicional: $r<1$, donde r es una coordenada radial medida desde el origen. En este sistema de coordenadas, la métrica sobre un espacio hiperbólico bidimensional es:

$$g_{ij} = \begin{pmatrix} \dfrac{4}{\left(1-x^2-y^2\right)^2} & 0 \\ 0 & \dfrac{4}{\left(1-x^2-y^2\right)^2} \end{pmatrix} = \begin{pmatrix} \dfrac{4}{\left(1-r^2\right)^2} & 0 \\ 0 & \dfrac{4}{\left(1-r^2\right)^2} \end{pmatrix} \qquad (7.14)$$

Aquí han pasado algunas cosas. Las coordenadas reales son (x, y), pero en la segunda expresión hemos escrito los elementos métricos en referencia a r = $\sqrt{x^2+y^2}$. No hay ningún problema, simplemente tenemos que recordar que r es una función de x e y. Más significativo es el hecho de que el plano hiperbólico es infinito en tamaño —tiene forma de silla de montar y se extiende infinitamente en todas direcciones—, pero en las coordenadas que hemos utilizado el valor que puede alcanzar r es finito. Esto es perfectamente correcto. Recordemos, si nos retrotraemos hasta el capítulo 2, que podemos trasponer una longitud infinita a un intervalo finito y viceversa. Es lo que sucede aquí. El disco de Poincaré es de tamaño infinito, pero lo hemos comprendido en coordenadas que se extienden tan solo en un intervalo finito.

De hecho, podríamos decir que realmente el espacio es infinitamente grande simplemente observando la métrica (7.14). Piense en lo que sucede a medida que nos aproximamos al borde del disco, $r \to 1$. Ambos elementos métricos distintos de cero contienen un factor $1/(1-r^2)^2$. A medida que r se acerca cada vez más a 1, $(1-r^2)$ se acerca cada vez más a 0, y así $1/(1-r^2)^2$ se expande hasta el infinito. Físicamente, esto significa que todo incremento de coordenada dx o dy se corresponde con una distancia real cada vez mayor. Aunque el alcance de las coordenadas no pase de un intervalo finito, el tamaño físico de la variedad descrita es infinitamente grande. Es la magia de la métrica que nos ocupa.



La figura de arriba muestra una representación de espacio hiperbólico en coordenadas de disco de Poincaré, en forma de mosaico de teselas triangulares. Parece como si los triángulos se hicieran cada vez más pequeños a medida que se aproximan al borde, pero es tan solo una ilusión causada por nuestro sistema de coordenadas. En la métrica (7.14), todos los triángulos tienen el mismo tamaño, y en realidad la misma forma. Hay un número infinito de ellos, se apretujan infinitamente en aproximación al borde del disco. Es famosa la serie de grabados del artista neerlandés M. C. Escher, a la que llamó Límite circular, basada en la geometría representada en esta figura.

TENSORES

La métrica de una variedad nos dice cómo calcular distancias. Nos referíamos a esto como «tensor» métrico, pero no explicábamos qué era un tensor en términos más generales. ¿Qué sucede entonces?

La idea de una función sobre una variedad es bastante simple. Se establece una relación que va de los puntos de una variedad a los números reales: se asigna un número (el valor de la función en tal punto dado) a cada punto. Podemos utilizar una función para responder a una pregunta como: «¿Cuál es la densidad de la materia en cada uno de los puntos?». Estamos también familiarizados con los vectores, que además de magnitud tienen dirección. Los vectores responden a cuestiones como: «¿Cuál es la velocidad de una partícula en esta trayectoria en particular?».

A veces queremos responder a preguntas más complicadas, que pueden involucrar a múltiples vectores o direcciones en el espacio. Quizá nos interese saber: «¿En qué medida se superponen los vectores y ?». O bien: «¿Cómo se enroscará un conjunto de trayectorias inicialmente paralelas en sus desplazamientos a través de un espacio curvo?». Los tensores son cantidades geométricas que contienen la información necesaria para abordar

cuestiones de mayor complejidad, como las anteriores. Las funciones y los vectores son tipos de tensores, pero para pensar de forma productiva sobre el espacio-tiempo curvo, vamos a necesitar versiones algo más elaboradas.

Hay dos formas de considerar los tensores, cada una de las cuales es útil según la circunstancia. Una sería la forma en que hemos venido utilizándolos: como tablas de elementos, cada uno de los cuales está etiquetado mediante un conjunto de índices. Todas estas presentaciones de la métrica en forma de matriz son ejemplos de tensores, en este sentido. Hay requisitos muy puntillosos en cuanto al modo en que cambian los elementos cuando cambiamos coordenadas, pero no tenemos que preocuparnos por estos.

Los elementos de una matriz cuadrada tienen dos índices, uno para indicar la fila y el otro para indicar la columna. Pero los tensores no tienen por qué ser matrices cuadradas. Pueden tener un número cualquiera de índices, y los valores de los índices son siempre correlativos a las dimensiones de la variedad que estamos considerando.[3]

Un vector es simplemente un tensor con un índice. Pensamos en un vector como en una flecha con una longitud y una dirección, pero dado un sistema de coordenadas como (x, y, z), podemos expresar el vector en términos de sus elementos, constituidos por las proyecciones del vector a lo largo de cada eje:

$$v^i = \begin{pmatrix} v^x \\ v^y \\ v^z \end{pmatrix} \qquad (7.15)$$

De modo que un vector es simplemente un tensor con un único índice, mientras que la métrica es un tensor con dos índices. Una función ordinaria es un tensor con cero índices. Y puede haber más índices, lo cual hace difícil representar el tensor como una tabla de elementos, pero puede

hacerse, no hay más que intentarlo. Podemos considerar, por ejemplo, un tensor de tres índices como un vector de tensores de dos índices:

$$T^{ijk} = \begin{pmatrix} T^{1jk} \\ T^{2jk} \\ T^{3jk} \end{pmatrix} = \left(\begin{pmatrix} T & T & T \\ T^{121} & T^{122} & T^{123} \\ T^{131} & T^{132} & T^{133} \end{pmatrix} \begin{pmatrix} T^{211} & T^{212} & T^{213} \\ T^{221} & T^{222} & T^{223} \\ T^{231} & T^{232} & T^{233} \end{pmatrix} \begin{pmatrix} T^{311} & T^{312} & T^{313} \\ T^{321} & T^{322} & T^{323} \\ T^{331} & T^{332} & T^{333} \end{pmatrix} \right) \tag{7.16}$$

No veo por qué iba usted a querer hacer esto, pero puede hacerlo. Una vez tiene más de dos índices, es más fácil pensar simplemente en los elementos individuales, en lugar de expresar el tensor como una gran tabla.

La otra forma de considerar los tensores es como una correspondencia entre un conjunto de tensores y otro tensor. Sí, suena a razonamiento circular, pero al final todo se sostiene. Si tiene, por ejemplo, dos vectores vi y wi, puede extraer cierto número de ellos utilizando la métrica. Podría imaginar el tensor métrico como una caja negra, en la que introdujera dos vectores y de ella salieran más vectores.

El número en cuestión se obtiene haciendo coincidir los índices de la métrica con los índices de los vectores, y sumando luego cada índice que se repite tanto arriba como abajo.

$$g(v,w) = \sum_{ij} g_{ij} v^i w^j = g_{11} v^1 w^1 + g_{12} v^1 w^2 + g_{21} v^2 w^1 + \cdots \tag{7.17}$$

Se trata de una elaboración famosa, al menos entre personas que andan con vectores de un lado para otro: es el producto interno, o «producto punto», entre los dos vectores:

$$\vec{v} \cdot \vec{w} = \sum_{ij} g_{ij} v^i w^j \tag{7.18}$$

En el espacio euclidiano ordinario, el producto punto de dos vectores es el producto de la longitud de ambos vectores por el coseno del ángulo que forman. Cuando los vectores apuntan en la misma dirección, el producto punto es simplemente el producto de sus longitudes, y cuando son perpendiculares el producto punto es siempre cero.

De este modo se revela un pequeñísimo secreto: la métrica no solo nos brinda la longitud de las curvas, también define lo que queremos decir con «perpendicular». Dos líneas que se cortan en un punto de intersección son perpendiculares cuando se anula el producto interno de dos vectores que

señalan a ese punto. Podemos comenzar a ver cómo la métrica determina todo lo que necesitamos saber acerca de la geometría de un espacio.

Habrá observado que unas veces anotamos los índices en forma de superíndices (como en el caso de los vectores y las coordenadas) y otras a modo de subíndices (como en el caso de la métrica). No se trata de una elección caprichosa. Las diferencias entre superíndices y subíndices son importantes en diferentes órdenes. Todo lo que necesitamos saber por ahora es que cuando sumamos índices entre sí, como en (7.17), solo podemos hacerlo cuando uno de los índices sumados está arriba y el otro abajo. Un índice que forma parte de la suma se llama índice mudo, mientras que un índice que no se suma es un índice libre. Los índices libres pueden tomar cualquier valor —siempre que un mismo índice tome el mismo valor en todos los términos de una ecuación—, mientras que los índices mudos no tienen un «valor», vienen a ser meras abreviaturas con la denotación: «súmese cualquier posible valor de este índice».

Sumar índices mudos es algo que se da tan a menudo en el análisis de tensores, que Albert Einstein inventó una estrategia útil, conocida como convenio de suma de Einstein. Siempre que tengamos un tensor o producto de tensores en los cuales se repita uno o más índices, tanto arriba como abajo, podemos obviar el signo explícito de sumatorio y asumir que los índices se suman. Tenemos así, por ejemplo:

$$g_{ij}v^i w^j = \sum_{ij} g_{ij}v^i w^j \qquad (7.19)$$

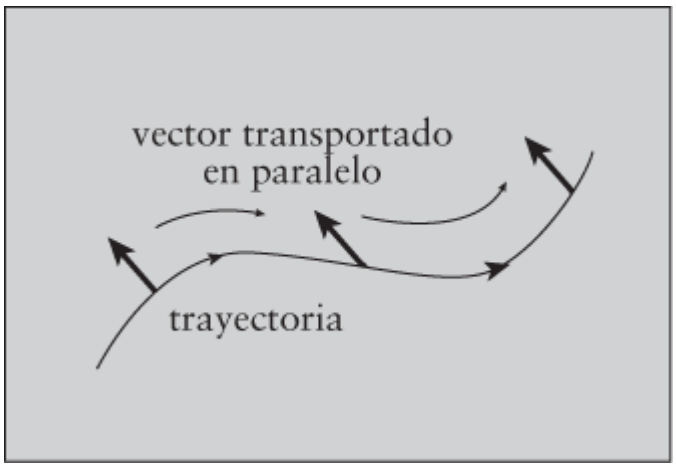
Al idear esta innovación, Einstein bromeó con un amigo: «¡He hecho un gran descubrimiento en matemáticas!». Sumar índices que se repiten no parece gran cosa, como concepto, pero adoptar esta convención tan útil ahorra mucho tiempo en una cuestión como la relatividad general.

TRANSPORTE PARALELO

La gran concepción de Riemann fue que la métrica de una variedad contiene toda la información que se necesita para obtener su curvatura y cualquier otra información geométrica que pueda interesarnos. Pensemos en cómo podemos hacer para extraer esa información. El primer paso será desplazar vectores de un lado para otro, porque el modo en que se desplacen se verá afectado por la curvatura subyacente. Para el resto de este capítulo, la maquinaria matemática se volvería engorrosa, de modo que vamos a destacar algunos puntos relevantes, mientras que dejaremos los detalles para el Apéndice B.

Imagine que está parado en algún lugar dentro de una variedad curva y sosteniendo un vector. Pongamos que tiene un giroscopio giratorio cuyo eje apunta en alguna dirección espacial. Y hay un amigo suyo a cierta distancia, sosteniendo también un vector. A usted le gustaría comparar los dos vectores: saber si apuntan a la misma dirección o a dirección diferentes, si uno es más largo que el otro, etcétera. ¿Cómo podemos hacerlo?

En un espacio plano —del que proceden la mayor parte de nuestras intuiciones—, esto parecería una cuestión un poco tonta. No tendría más que caminar hacia su amigo, llevando su vector consigo y, fijando los vectores, compararlos al llegar a su posición. Pero, ¿qué significa «fijar los vectores»? Una respuesta sería establecer un sistema tradicional de coordenadas cartesianas y mantener constantes todos los componentes de nuestro vector. Ello nos permitiría desplazarlo de un lugar a otro, no hay problema.



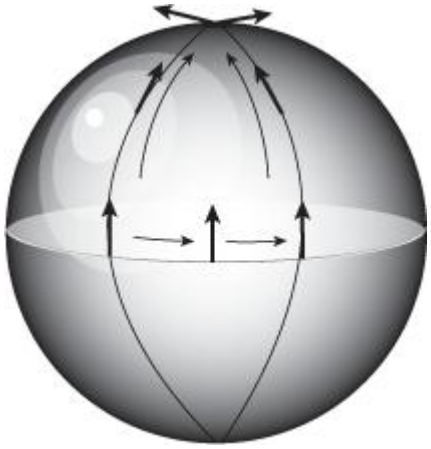
El problema está en que este procedimiento no funciona en geometrías no planas. No contamos con «coordenadas cartesianas», porque la existencia de tales coordenadas se basa en una métrica plana. Aunque tal vez esto no sea más que una complicación técnica, y de alguna manera podamos realizar una acción moralmente equivalente a mantener un vector constante mientras lo movemos.

Podemos. El transporte paralelo es el proceso por el que podemos partir con un vector definido en algún punto y desplazarlo por una trayectoria especificada, procurando en cada paso mantenerlo lo más paralelo posible a su orientación de partida. (Como puede suponer, los pasos tendrán que ser infinitesimalmente pequeños, así que recurriremos al cálculo para mover el vector de un punto a otro).

Surge, sin embargo, una matización sutil. En un espacio plano, no es solo que el sentido que demos a «mantener constante el vector» parezca intuitivo, sino que además no importa qué trayectoria sigamos para desplazar el vector de un lugar a otro. En un espacio curvo general esto no es así, como podemos apreciar si pensamos en el transporte paralelo sobre una esfera bidimensional.

Partamos de un punto en el ecuador, con un vector señalando hacia el norte. Caminemos en dirección al polo norte, haciendo todo lo posible por mantener el vector constante; no es muy difícil, ya que el vector tiene que apuntar en todo momento en la misma dirección en que nos desplazamos.

Imaginemos ahora que, en lugar de ir hacia el norte, caminamos durante un trecho a lo largo de la línea del ecuador, manteniendo igualmente el vector constante. En este caso, tenemos que llevar el vector de tal modo que siga apuntando hacia el norte. Tras recorrer cierta distancia, damos un giro de noventa grados y continuamos caminando en dirección norte, con nuestro vector.



Estas dos rutas diferentes para ir de nuestro punto de partida en el ecuador al polo norte, nos han dejado con sendos vectores apuntando en direcciones obviamente diferentes, a pesar de haber hecho todo lo posible por mantener los vectores contantes durante todo el tiempo. Tal cosa no habría sucedido en un espacio plano, pero es inevitable en una esfera, sobre la cual el transporte paralelo siguiendo trayectorias diferentes conduce generalmente a resultados diferentes. Como veremos en un momento, este fracaso a la hora de intentar que dos caminos nos den el mismo resultado nos proporciona una manera de formalizar aquello que entendemos por «curvatura». (Tenga en cuenta que esta operación es enteramente intrínseca a la esfera, no tenemos que observarla desde el exterior para transportar un vector en paralelo).

Hemos tropezado con una característica profunda, y a veces difícil de interiorizar, de espacio curvo (o espacio-tiempo): no hay una única manera de comparar vectores situados en puntos diferentes. Podemos transportar un vector en paralelo a lo largo de una determinada curva, pero si siguiéramos una curva diferente, podríamos obtener un resultado diferente. Esto

significa, por ejemplo, que no podemos hablar en realidad de «la velocidad» de una galaxia lejana en un universo en expansión. De hecho hablamos continuamente de ello, pero implícitamente estamos tomando determinadas opciones arbitrarias acerca de cómo comparar vectores. Lo cual está muy bien, pero siempre que no dejemos de tener presente qué cosas son únicas y están bien definidas, y cuáles simplemente parecen las más adecuadas. Nos encontramos aquí con una versión readaptada de la enseñanza que ya predicamos en el contexto del experimento mental de los gemelos, en el capítulo 6: es mejor pensar en términos locales y comparar cantidades definidas en el mismo punto, que autoengañarnos comparando cosas que suceden lejos de nosotros.

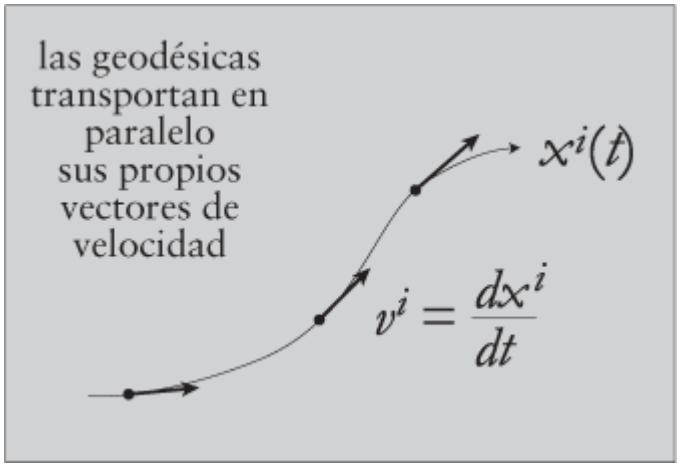
GEODÉSICAS

Al principio del Capítulo 3, mencionamos dos formas de diseñar una línea recta entre dos árboles. Una de ellas consistía en disponer una cuerda y tensarla al máximo —una manera muy tangible de construir un camino lo más corto posible— y la otra en caminar simplemente sin abandonar una dirección fija. Ambos métodos producían el mismo resultado. Y lo mismo sigue sucediendo en el caso de una variedad curva general en la geometría de Riemann, si bien, cuando el espacio es curvo, puede que no parezca muy natural concebir la trayectoria resultante como «recta». En una esfera, por ejemplo, tales trayectorias forman circunferencias o arcos de circunferencia.

La trayectoria que minimiza la distancia entre dos puntos (o que maximiza el tiempo propio, en términos de espacio-tiempo, cuando lo veamos) se conoce como geodésica. Las líneas geodésicas obedecen a una ecuación (Apéndice B) que puede derivarse de un modo muy similar a cómo lo hacíamos al hablar del principio de mínima acción, en el Capítulo 3. Allí, considerábamos cada una de las trayectorias que podía seguir una partícula, asociábamos una acción a cada posibilidad y pedíamos que la acción tuviera una derivada cero en el espacio de trayectorias posibles. El proceso de encontrar líneas geodésicas es análogo, salvo que lo que minimizamos es la

longitud de una curva, en lugar de la acción de la trayectoria de una partícula.

Además de constituir trayectorias lo más cortas posibles, las geodésicas son igualmente resultantes de caminar en línea recta. La versión formal de esta idea se sirve del transporte paralelo. Consideremos una trayectoria, especificada por la secuencia de puntos que conforman el camino, y también un parámetro que nos dice en qué punto nos encontramos a lo largo de la trayectoria. Podríamos escribir xi(t), por ejemplo, donde xi son las coordenadas en cualesquiera sean las dimensiones en las que estemos, y t es el parámetro a lo largo de la trayectoria. (A menudo el parámetro es «tiempo», en realidad, pero aquí solo usaremos t como una variable conveniente). Tenemos entonces un vector, la velocidad vi = dxi/dt, que apunta hacia la dirección en la que nos desplazamos, con una longitud que es la que corresponde a nuestra celeridad.



«Seguir desplazándonos en la misma dirección» significa mantener nuestro vector velocidad tan constante como sea posible mientras nos desplazamos. Esto es exactamente lo que hace el transporte paralelo. Así pues, otra forma de definir una geodésica sería: «una trayectoria cuya velocidad en cada punto es la obtenida al transportar en paralelo el vector velocidad inicial». Así es, a fin de cuentas, como el proceso del transporte paralelo se relaciona con el tensor métrico: las curvas que transportan en paralelo su propio vector velocidad tendrán que ser también las curvas con la mínima longitud posible.

CURVATURA

En este punto nos encontramos: el tensor métrico es la estructura geométrica más básica que puede tener una variedad. Nos permite calcular la distancia que recorre una trayectoria, áreas y volúmenes de regiones del espacio de dimensiones superiores, así como el producto interno entre vectores. También nos dice cómo transportar en paralelo vectores a lo largo de una curva. Esto nos proporciona un buen modo de abordar las líneas geodésicas, como curvas que transportan en paralelo sus propias velocidades, además de constituir la distancia más corta para una trayectoria. Pero el transporte paralelo resulta también crucial para encajar

la última pieza del puzle: una caracterización completa de los tipos de curvatura que puede adoptar un espacio.

Tanto la esfera como el plano hiperbólico son variedades curvas, pero su curvatura es la más simple que podríamos concebir: una curvatura que conserva la misma apariencia en todos los puntos y en cualquier dirección. Nos gustaría ser capaces de formular una sólida caracterización local del grado de curvatura que puede tener una variedad en cada uno de sus puntos. La métrica, como hemos visto, por sí misma no es perfecta para el cometido, puesto que puede antojársenos simple o complicada dependiendo del sistema de coordenadas, aunque la geometría subyacente sea la misma. Nos gustaría definir una cantidad —algún tipo de tensor, por ejemplo—, que nos revelara de inmediato el grado de curvatura de un espacio, dando el valor cero cuando el espacio es plano, y diferente de cero cuando el espacio es curvo.

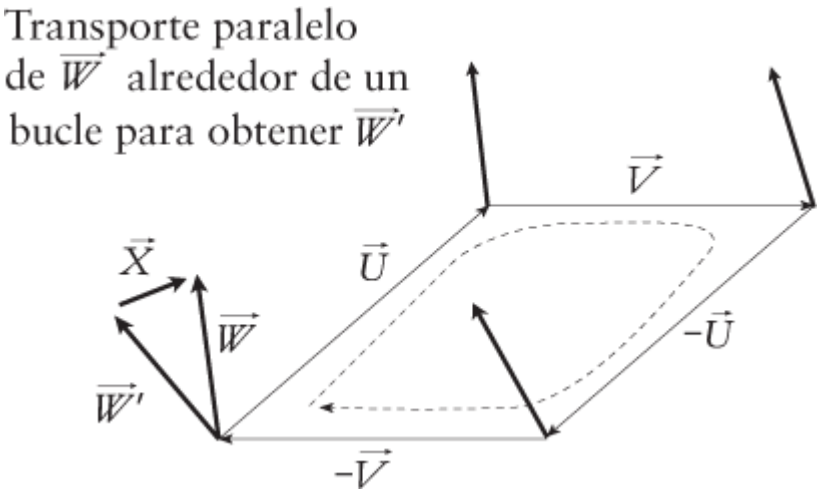
Cuando transportamos en paralelo un vector a lo largo de dos trayectorias diferentes sobre la superficie de una esfera, terminamos con dos vectores diferentes en el polo norte. De un modo equivalente, podríamos haber partido del polo norte, haber seguido nuestro camino hasta el ecuador, haber recurrido una porción de este y regresado luego, desplazándonos en bucle. La idea crucial es esta: si transportamos en paralelo un vector siguiendo una trayectoria en forma de bucle, al regresar al punto de partida no estará necesariamente alineado con el mismo vector en el momento de partir. Al menos, no si el espacio a través del cual se desplaza es curvo.

Esto sugiere una manera de caracterizar la existencia de curvatura. En un espacio plano, el transporte paralelo siguiendo un bucle cerrado nos dará siempre el mismo vector de partida. En una variedad curva, el vector puede rotar a lo largo del camino.

El problema está en que hay demasiados bucles cerrados. Nadie podría formular fácilmente los modos en que los vectores cambian al ser transportados por todas las trayectorias cerradas concebibles. Lo que nos interesa es elegir un conjunto especial de bucles, aquellos que podamos caracterizar fácilmente con apenas unos cuantos números.

Podemos abordar este problema mediante la socorrida estrategia que hasta el momento nos está resultando tan práctica: utilizando el cálculo infinitesimal para alcanzar resultados de tamaño finito a voluntad. Por este motivo, el estudio de curvaturas arbitrarias sobre variedades se llama geometría diferencial.

Veamos cómo generar diminutos bucles para movernos por curvas. Tomemos dos vectores, y , ambos definidos en el mismo punto p. Partiendo de p, nos desplazamos una distancia infinitesimal siguiendo la dirección de . Desde ahí, nos desplazamos una distancia infinitesimal siguiendo la dirección de . (Técnicamente, la distancia que cubrimos en cada desplazamiento es proporcional a la longitud de y). A continuación retrocedemos desde el punto al que hemos llegado, siguiendo la dirección opuesta a , y finalmente volvemos a retroceder siguiendo la dirección opuesta a . Habremos regresado así a nuestro punto de partida p, tras describir un diminuto paralelogramo en el transcurso de nuestro viaje.[4]



Es este un buen modo de definir un pequeño bucle utilizando una cantidad de información relativamente compacta: únicamente los componentes de dos vectores en un punto. Para, partiendo de ahí, llegar a medir una curvatura, consideremos un tercer vector en el mismo punto original de partida. Transportamos ahora en paralelo alrededor de nuestro diminuto paralelogramo, lo cual resultará en un nuevo vector ’. En una variedad plana, el vector no cambiaría, y tendríamos ’ = . Pero en una variedad curva

puede haberse modificado, aunque sea muy poco. Podemos definir la diferencia en forma de un nuevo vector más:

$$\vec{X} = \vec{W} - \vec{W}' \qquad (7.20)$$

Así es como especificamos la curvatura en cualquier punto de una variedad arbitraria. Utilizamos dos vectores para definir un bucle, transportamos un tercer vector en paralelo alrededor del bucle, y a partir de ahí un cuarto vector que representa el cambio sufrido por el tercero constituirá una medida del grado de curvatura del espacio. Si es casi plano, será pequeño; en un espacio muy curvo, será relativamente grande.

Lo que tenemos, en otras palabras, es una correspondencia que va de un conjunto de tres vectores a un cuarto vector, . Pero ya tenemos un nombre para correspondencias entre agrupaciones de vectores: son tensores. Así pues, lo que tenemos en realidad es el tensor de curvatura de Riemann, el cual introduce como datos dos vectores para definir un bucle y un tercer vector para ser transportado, y obtiene como resultado un cuarto vector para caracterizar el cambio alrededor del bucle.

Sigue pudiendo parecer un poco engorroso preguntarnos qué sucede con todos y cada uno de los vectores singulares que rodean cada paralelogramo definido por otros dos vectores. Aquí es donde la magia de los componentes viene para salvarnos. Pensemos en cada uno de nuestros vectores en términos de sus componentes, de tal forma que representaremos por un conjunto de componentes Ui, y lo mismo para los demás vectores. El

número de componentes de cada vector es equivalente a la dimensionalidad de la variedad en que estamos.

Así, al igual que la información acerca del elemento de línea puede cifrarse en los componentes gij del tensor métrico, el tensor de curvatura de Riemann puede también cifrarse en un conjunto de componentes, expresados como Rijkl. (El orden de los índices importa, como también si están arriba o abajo).

Esto nos proporciona una manera explícita de representar la correspondencia de a , utilizando componentes:

$$X^i = R^i{}_{jkl} W^j U^k V^l \qquad (7.21)$$

Advierta el convenio de suma de Einstein en acción. En la parte de la derecha, estamos sumando todos los valores de los índices que se repiten, j, k y l.

Hay que reconocer que hay un montón de componentes en Rijkl. Cuatro índices, cada uno de los cuales toma un número d de valores en una variedad d-dimensional, lo cual nos da d^4 componentes en total. Son 81 componentes para tres dimensiones, 256 componentes para cuatro dimensiones, y el número se incrementa exponencialmente, en sentido literal.

Por fortuna hay simetrías del tensor de Riemann que facilita un poco las cosas a los profesionales de la relatividad general. Los diferentes componentes se relacionan entre sí de diversas formas, más simples. Así, por ejemplo, si se permutan los dos últimos índices se obtiene el mismo número, pero con un signo menos delante: Rijkl = Rijlk. Gracias a esta y otras relaciones, el número total de componentes independientes en d dimensiones resulta ser $^1/12\ d^2(d^2 1)$. Para cuatro dimensiones, esto son veinte componentes independientes, un número considerable pero manejable. Para dos dimensiones no es más que un componente. Esto significa que en una variedad bidimensional, hay un único número en cada

punto que caracteriza el grado de curvatura de la superficie (positiva o negativamente, como sabemos). Para una dimensión, ¡hay cero componentes independientes! Pensamos que podemos trazar una curva de una dimensión que parezca curva, pero ello se debe una vez más a nuestra perspectiva intrínseca que nos engaña. En el interior de un mundo de una dimensión, no existiría nada que respondiera a la noción de curvatura.

La curvatura depende de la métrica en un sentido bien definido, pero esta dependencia es compleja, por lo que, también en esta ocasión, hemos dejado los detalles más arduos para el Apéndice B. Lo que importa es que entienda la base conceptual de lo que hemos hecho: definir una métrica, luego el transporte paralelo y las líneas geodésicas, y finalmente la curvatura. Estas serán las herramientas matemáticas necesarias para dar sentido a la relatividad general de Einstein.

1 A partir del debate público, uno puede tener la impresión de que se da una especie de jerarquía, desde la mera «hipótesis» (que vendría a ser poco más que una suposición), pasando por un «modelo» y una «teoría», hasta finalmente una «ley»; proceso durante el cual iría estableciéndose un listado de ideas en orden de credibilidad creciente. En la práctica entre los científicos reales, sin embargo, los términos se solapan considerablemente, hasta el punto de no existir una distinción útil entre ninguno de tales términos. Hay, por otro lado, una distinción crucial y ampliamente reconocida entre «teoría», que constituye un modelo científico acerca de cómo funciona el mundo, y «teorema», que representa un resultado matemático rigurosamente demostrado.

2 Utilizamos «matriz» (en inglés, matrix) en el sentido matemático de «una tabla de cantidades». Nada que ver con la eventualidad de estar viviendo o no en una realidad simulada por ordenador.

3 La métrica y otros tensores que estemos considerando pueden tener diferentes valores para cada punto del espacio (o del espacio-tiempo, cuando procedamos a ello). Así que de lo que estamos hablando realmente es de campos tensoriales. Un «campo escalar» es simplemente una función, que toma un valor numérico en cada punto; tenemos también «campos vectoriales» y otros campos tensoriales.

4 Si teme que la curvatura nos lleve a un punto ligerísimamente desplazado con respecto a nuestro punto de partida, tiene razón. Pero nuestro paralelogramo es lo bastante pequeño como para poder despreciar el desajuste por comparación con las cantidades que nos interesan en este momento.

OCHO

GRAVEDAD

La mecánica clásica, cualquiera que sea la impresión que haya podido recibir a través de los capítulos precedentes, no es una teoría de la física. Es más bien un marco conceptual en el seno del cual pueden elaborarse teorías legítimas. La mecánica clásica dice: «Un sistema físico se describe mediante posiciones y momentos o generalizaciones apropiadas a partir de ellos, y sus variables obedecen a ecuaciones formuladas de forma equivalente a (Newton), o a las ecuaciones de Hamilton, o al principio de mínima acción». Permanece agnóstica con respecto a qué sean las fuerzas, el hamiltoniano o la acción. La gravedad newtoniana, en cambio, es una teoría específica; proporciona reglas precisas acerca de qué son las fuerzas (ley de la inversa del cuadrado), a partir de las cuales pueden hacerse predicciones inequívocas. Hay muchas teorías específicas que tienen cabida dentro del marco conceptual de la mecánica clásica.

La mecánica clásica es un marco conceptual alternativo a la mecánica clásica. Se da una teoría específica del oscilador armónico simple clásico, y una teoría separada del oscilador armónico simple cuántico, por ejemplo. Casi todas las teorías concebidas por los físicos modernos entran dentro de alguno de estos dos marcos conceptuales, el «clásico» o el «cuántico».

También la relatividad es un marco conceptual, no una teoría. Las teorías específicas pueden ser «no relativistas» (como la gravedad newtoniana) o «relativistas» (como la teoría del electromagnetismo de Maxwell). La distinción atañe a nuestro modo de pensar acerca del espacio-tiempo: las teorías no relativistas postulan un espacio y un tiempo absolutos y la acción instantánea a distancia, mientras que las teorías relativistas cuentan con conos de luz y un firme límite para la velocidad. Las categorías relativista/no relativista se entrecruzan con las categorías clásico/cuántico; hay modelos que encajan en cualquiera de las categorías correspondientes.

	CLÁSICO	CUÁNTICO
No relativista	Gravedad newtoniana	Oscilador armónico cuántico
Relativista	Electromagnetismo de Maxwell	Electrodinámica cuántica

En física, sucede con frecuencia que llegamos a comprender cierto fenómeno ya sea en un contexto clásico o no relativista, y luego necesitamos trabajar un poco para llegar a una comprensión análoga en un marco cuántico o relativista.

Las ideas de la relatividad se inspiraron ampliamente en las características del electromagnetismo de Maxwell, que ha quedado como paradigma de teoría relativista clásica. Pero una vez la relatividad había entrado en juego, era natural volverse hacia las demás fuerzas de la naturaleza y preguntarse cómo podían encajar en el nuevo marco conceptual. El ejemplo obvio a considerar era la gravedad, que Newton había descrito con tanto éxito en el nacimiento mismo de la mecánica clásica no relativista.

Conceptualizar con el mismo éxito una teoría relativista de la gravedad resultó más difícil de lo que nadie había previsto. Tuvo que pasar una década entera entre los artículos de Einstein sobre la relatividad especial en 1905 y su formulación final de la relatividad general en 1915. Y no era el único que trabajaba sobre la cuestión. El físico finlandés Gunnar Nordström desarrolló una teoría que entraba en competencia, y él y Einstein mantuvieron correspondencia mientras trabajaban. Pero la capacidad de penetrar en las cuestiones físicas de Einstein, así como sus dotes para el raciocinio en torno a los experimentos mentales terminaron por imponerse. Sus contribuciones a la relatividad especial fueron enormemente importantes, y fue su trabajo fundacional sobre mecánica cuántica (también en 1905) lo que le valió el Premio Nobel, pero la propuesta de Einstein de la relatividad general como una teoría del espacio-tiempo curvo fue lo que lo convirtió en el científico más célebre del siglo XX.

MASA INERCIAL Y GRAVITACIONAL

La masa y la gravedad tienen una relación especial. La segunda ley de Newton dice que la fuerza que actúa en un cuerpo es igual al producto de su masa por la aceleración que experimenta. A su vez, la ley gravitatoria de la

inversa del cuadrado de Newton dice que la fuerza gravitatoria entre dos objetos es proporcional a la masa de cada uno de ellos, y a también a la inversa del cuadrado de la distancia.

La idea de «masa» aparece en ambas relaciones. Pero el papel desempeñado por esta idea es completamente diferente en un caso y en otro. En la segunda ley, la masa es una medida de la inercia: la resistencia que el objeto opone a ser acelerado. La cantidad es la misma, sea cual sea la fuerza que causa la aceleración. Pero en la ley de la gravedad, la masa mide la interacción entre el cuerpo y la fuerza de gravedad específica. ¿Por qué estas dos cantidades, conceptualmente distintas, son numéricamente iguales entre sí?

Para verlo con más detenimiento, sea m la masa inercial del cuerpo y M su masa gravitacional. La segunda ley es:

$$\vec{F} = m\vec{a} \qquad (8.1)$$

Mientras que la ley gravitatoria de la inversa del cuadrado para dos objetos, 1 y 2, se formula:

$$\vec{F} = G\frac{M_1 M_2}{r^2}\vec{e}_r \qquad (8.2)$$

Sin embargo, sucede que en el mundo real las masas inercial y gravitacional son iguales entre sí:

$$m = M \qquad (8.3)$$

No hay una razón obvia para que esto tenga que ser así. Consideremos la fuerza entre dos partículas con carga eléctrica, tal como nos la proporciona la ley de Coulomb. También adopta una forma de inversa del cuadrado, así:

$$\vec{F} = K\frac{Q_1 Q_2}{r^2}\vec{e}_r \qquad (8.4)$$

Tiene una apariencia muy similar a la ley de gravitación de Newton, con la constante de Coulomb K reemplazando a la constante de Newton G, y las cargas eléctricas Q1 y Q2 sustituyendo a las masas gravitacionales. Pero hay una diferencia crucial. En contraste con la gravedad, en el electromagnetismo la carga que actúa como una fuente para la fuerza eléctrica no tiene nada que ver con la masa inercial del objeto. Puede ser positiva (como en un protón), negativa (como en un electrón), o incluso cero (como en un neutrón). Las diferentes partículas responden a la fuerza electromagnética de forma también diferente, y algunas ni siquiera responden. En cambio, toda partícula responde a la gravedad, de un modo en proporción exacta a su masa, y la gravedad siempre ejerce atracción.

La teoría de la gravedad de Newton no intenta explicar por qué la masa que es origen de la gravedad es la misma que la masa que rige la inercia y el movimiento. Simplemente lo tomamos como un hecho que se da en el universo y seguimos adelante.

EL PRINCIPIO DE EQUIVALENCIA

La equivalencia entre la masa inercial y la masa gravitacional tiene una sugestiva consecuencia. Consideremos dos partículas que se atraen una a la otra por acción de la gravedad, y pensemos en lo que le sucede a la partícula 2 por influencia de la partícula 1. Combinando (8.1) y (8.2) tenemos:

$$\vec{F} = m_2\vec{a} = G\frac{M_1 M_2}{r^2}\vec{e}_r \qquad (8.5)$$

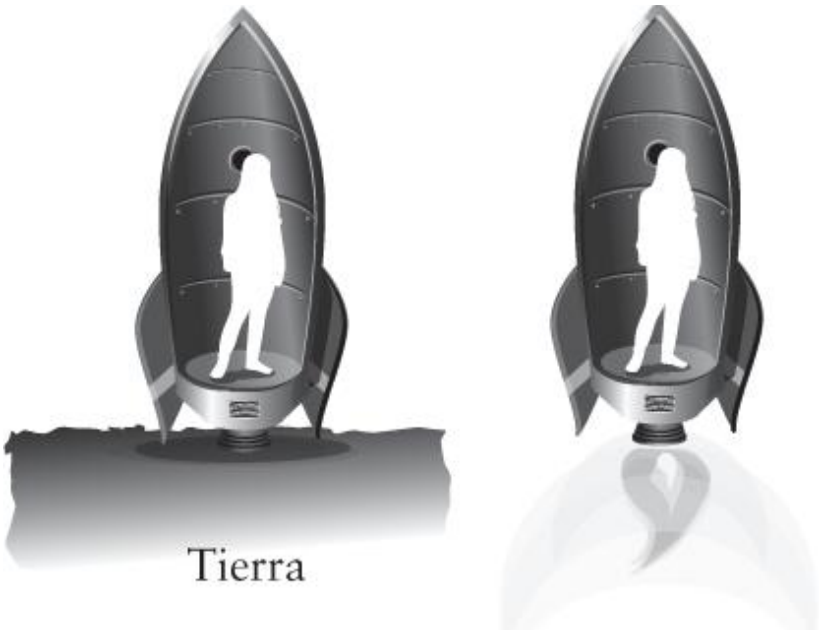
Pero m2 = M2, de modo que estas dos cantidades se anulan entre sí, y queda:

$$\vec{a} = G\frac{M_1}{r^2}\vec{e}_r \qquad (8.6)$$

Esta fórmula expresa la aceleración experimentada por la partícula 2, que resulta ser completamente independiente de la masa de la partícula 2. Dicho de otro modo, todo objeto experimenta una aceleración similar en un campo gravitatorio, sea cual sea su masa (o cualquier otra cosa).

Esto es interesante. Ciertamente, no vale para otras fuerzas, como el electromagnetismo. Dado un campo eléctrico, una partícula con carga positiva se verá impelida en una dirección y una partícula con carga negativa en la dirección opuesta. La gravedad, por el contrario, lo atrae todo en la misma dirección y con la misma aceleración. Esta percepción es anterior a las ecuaciones de Newton, se remonta a Galileo.

El gran avance de Einstein fue tomar esta característica de la gravedad — todas las partículas caen con la misma aceleración en un campo gravitatorio — y hacerla extensiva en forma de principio general. En el Capítulo 3 observábamos que si estuviéramos encerrados en un cohete espacial, notaríamos si acelerábamos o no acelerábamos, pero no habría manera de estimar nuestra velocidad absoluta. Comparemos ahora un cohete estacionado sobre el suelo con otro que está acelerando por el espacio a una g (la aceleración de la gravedad sobre la superficie de la Tierra).



En el cohete con aceleración, si soltamos un objeto, este parece caer al suelo debido al movimiento del cohete. Todos los objetos caerán de modo idéntico, aunque en realidad sea el cohete el culpable. Pero, a causa de la igualdad entre las masas inercial y gravitacional, lo mismo sucederá en el cohete estacionario sobre el suelo. Cualquier objeto que soltemos caerá con la misma aceleración, sea cual sea su masa.

Esto llevó a Einstein a proponer el principio de equivalencia: en áreas pequeñas de espacio-tiempo, los efectos de la gravedad son equivalentes a los experimentados en un marco de referencia acelerado. No es simplemente cuestión de soltar objetos para ver cómo caen; podemos hacer todos los experimentos que queramos en el interior de nuestra nave espacial cerrada, y nos dirán si estamos acelerando a través del espacio o posados apaciblemente en el suelo. En términos de operatividad: para áreas pequeñas del espacio-tiempo, las leyes de la física se reducen a las de la relatividad especial (no gravitatoria).

Cuando consideramos áreas más grandes del espacio-tiempo, esto no funciona, claramente. El campo gravitatorio puede ejercer su fuerza de atracción en diferentes direcciones, dependiendo de dónde estemos situados, mientras que un marco de referencia acelerado siempre conducirá a un movimiento uniforme para cualquier objeto ubicado en su interior. Cerca de

la Tierra, por ejemplo, la fuerza de gravedad siempre apuntará hacia el centro del planeta. Si nuestra nave espacial fuera de un tamaño similar al de la Tierra, podríamos apreciar que varios péndulos separados entre sí a una distancia considerable apuntarían en direcciones diferentes. De modo que la restricción del principio de equivalencia a áreas pequeñas del espacio-tiempo es crucial.

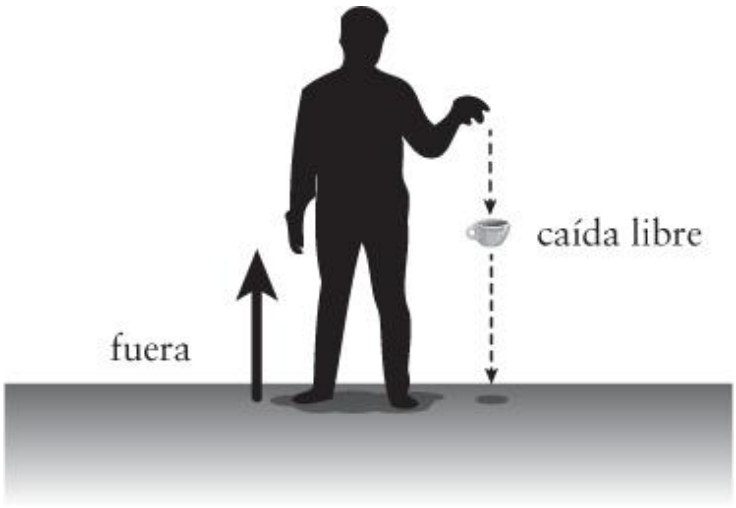
Usted o yo, habiendo dado con esta idea, podríamos sentirnos orgullosos de nosotros mismos y dar por concluida la jornada. Pero Einstein era Einstein, y dio un paso más. El hecho de que «la física adopta el aspecto de la relatividad especial para áreas pequeñas del espacio-tiempo» recuerda sospechosamente a algo que ya hemos oído antes. En el último capítulo, de hecho: Bernhard Riemann advirtió que el estudio de las variedades curvas le llevaba a postular que «la geometría parece euclidiana en áreas pequeñas del espacio». Así pues, razonó Einstein, quizá no deberíamos pensar en la gravedad como en una «fuerza» que reside en el interior del espacio-tiempo. Quizá deberíamos pensar en ella como en una característica del espacio tiempo mismo: la curvatura.

CAÍDA LIBRE

Parece absurdo, ¿verdad? ¡Por supuesto que la gravedad es una fuerza! Las manzanas caen de los árboles, las personas tropiezan y se caen por las escaleras, la Tierra gira en torno al sol. ¿Qué sentido tiene pensar en la gravedad si no es como en una fuerza?

Para ser sinceros, sigue pudiendo pensar en la gravedad como en una fuerza cuando tiene sentido hacerlo, a los físicos de partículas les gusta hablar de las cuatro fuerzas de la naturaleza, entre las cuales incluyen la gravedad junto con el electromagnetismo y las fuerzas nucleares fuerte y débil. Pero la gravedad es un tipo de fuerza diferente, precisamente porque es universal. La gravedad afecta a todo de un mismo modo, a diferencia de otras fuerzas, que afectan a los objetos de forma diversa dependiendo de su carga. Esto es lo que nos permite pasar de concebirla como una fuerza que se propaga en el seno del espacio-tiempo, a pensar en ella como en una propiedad del espacio-tiempo en sí mismo.

La concepción de Einstein requiere de un cambio de perspectiva. Pongamos que suelta usted una taza de café y que esta cae al suelo. Es tentador decir: «la fuerza de la gravedad ha tirado de ella hacia abajo». Pero de acuerdo con la relatividad general, el movimiento natural de las cosas es estar en caída libre. La taza de café no ha hecho más que lo que le es propio. Es usted, que permanece de pie sobre el suelo, quien está experimentando una fuerza: la Tierra le empuja hacia arriba presionando bajo las suelas de sus zapatos, y sustrayéndolo de una trayectoria en caída libre. Se trata de un mismo conjunto de sucesos, pero descritos de un modo diferente.



En la relatividad especial, los objetos que no experimentan fuerzas se desplazan en línea recta. Ya sabemos en qué consiste la generalización de una línea recta en una variedad curva: una línea geodésica. Según la relatividad general, el espacio-tiempo es curvo, y los objetos que no experimentan aceleración se desplazan a lo largo de líneas geodésicas en el seno de esa variedad curva.

Podría objetar con el hecho de que la Tierra se mueve alrededor del sol siguiendo una trayectoria próxima a una elipse, que a nadie sugiere la idea de lo que es una línea recta. Esto es porque sigue anclado en el hábito de pensar en el espacio, en lugar de pensar en el espacio-tiempo. En el marco de reposo del sol, la Tierra se mueve principalmente en una dirección de tipo temporal, ya que su desplazamiento a través del espacio es muy lento (alrededor de 0,0001 veces la velocidad de la luz). Pero la influencia gravitatoria del sol hace que el espacio-tiempo a su alrededor se curve levemente, por lo que una línea geodésica en torno a él tiene la apariencia de una elipse en el espacio. La Tierra hace todo lo posible por desplazarse en línea recta, exactamente igual que la taza de café.

LA MÉTRICA EN EL ESPACIO-TIEMPO

En el Capítulo 6 hablamos del espacio-tiempo de Minkowski y de cierta similitud entre las fórmulas del tiempo propio y la longitud de las curvas del espacio-tiempo con el teorema de Pitágoras, solo que con el curioso añadido de un signo menos. En el Capítulo 7 hablamos de la geometría de Riemann, cuya idea fundamental es la del tensor métrico, que nos proporciona un elemento de línea a partir del cual podemos calcular las longitudes de las curvas. Ambas cosas están estrechamente relacionadas. La relatividad se construye sobre la idea de que el espacio-tiempo tiene un tipo de métrica que le es propia, con un signo menos para la dirección de tipo temporal. Una métrica con este tipo de signo menos se conoce como lorentziana, y si el espacio-tiempo es perfectamente plano tenemos la métrica de Minkowski.

Recuerde que contamos con una linda estrategia para indexar dimensiones del espacio-tiempo, utilizando índices griegos para las cuatro dimensiones espacio-temporales ($x\mu$), y luego subdividiéndolos de forma que reservamos el cero para el tiempo ($x^0 = t$) y los índices latinos para el espacio ($xi = x^1, x^2, x^3$). Con arreglo a esta notación, la métrica de Minkowski adopta la forma:

$$g_{\mu\nu} = \begin{pmatrix} g_{00} & g_{01} & g_{02} & g_{03} \\ g_{10} & g_{11} & g_{12} & g_{13} \\ g_{20} & g_{21} & g_{22} & g_{23} \\ g_{30} & g_{31} & g_{32} & g_{33} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix} \quad (8.7)$$

En forma de elemento de línea, esto es equivalente a:

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \quad (8.8)$$

El signo menos genera cierta incomodidad, pero uno aprende a lidiar con ello. Tal como lo hemos expresado, puede extraerse la longitud de una curva de tipo espacial directamente a partir de la métrica, cosa que ya hicimos esencialmente en la ecuación (6.10). Pero para las curvas de tipo temporal —que, como sabe, son aquellas en que se mueven los objetos reales—, el

intervalo espacio-temporal ds^2 será negativo, por lo que $ds = \sqrt{ds^2}$ resulta un número imaginario, lo cual se ve raro. Está bien, para las curvas de tipo temporal trabajaremos en su lugar con el tiempo propio, que se define como:

$$d\tau^2 = -ds^2 \qquad (8.9)$$

Una alternativa sería multiplicar simplemente la parte derecha entera de (8.8) por −1, y definir así el elemento de línea como menos lo que hemos elegido que sea. Entonces, el elemento de línea nos da el tiempo propio directamente y hay que multiplicarlo poro −1 cuando calculamos distancias de tipo espacial. Es una opción respetable, resultado de elegir determinada convención, y numerosos manuales trabajan siguiéndola. (Como orientación, digamos que, grosso modo, los físicos especializados en relatividad utilizan en su mayor parte la convención −+++ que hemos elegido aquí, mientras que los físicos de partículas a menudo optan por +−++). Nuestra opción es conveniente si a usted le gusta de vez en cuando pensar en «el espacio en un momento del tiempo», que de esta manera hereda una buena métrica euclidiana +++.

Si quisiera reducir la relatividad especial hasta dejarla en una simple frase compacta, quedaría en algo así como: «el espacio-tiempo tiene una métrica de tipo Minkowski». Las ecuaciones (8.7) y (8.8) contienen toda la información necesaria para hablar de distancias, tiempos, conos de luz, viajes de gemelos (y otros aventureros) por el espacio, y otras cosas más.

La percepción de Einstein era que el espacio-tiempo real de nuestro universo se parece al espacio-tiempo de Minkowski en áreas pequeñas, pero que su tejido, entendido en su totalidad, está urdido de tal modo que lleva a la curvatura, tal como imaginara Riemann. La geometría entera puede describirse mediante algún tipo de métrica, pero no será algo tan sencillo como en (8.7). Resolver la métrica del espacio-tiempo y descifrar qué influencia ejerce en la materia del universo es aquello a lo que los profesionales de la relatividad general dedican la mayor parte de sus horas de trabajo.

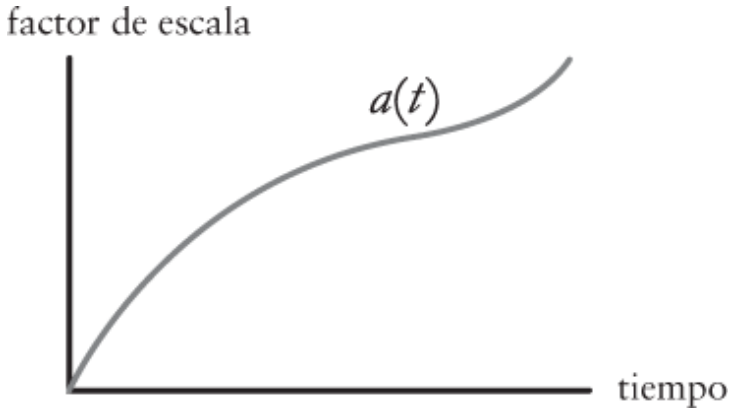
Observemos una métrica simple, algo diferente de la de Minkowski. Esta es la métrica para un universo en expansión:

$$g_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & a^2(t) & & \\ & & a^2(t) & \\ & & & a^2(t) \end{pmatrix} \qquad (8.10)$$

Si un elemento de la métrica es cero, dejamos simplemente el espacio en blanco, confiando en que usted lo rellenará mentalmente si fuera necesario. En forma de elemento de línea, sería:

$$ds^2 = -dt^2 + a^2(t)\left[dx^2 + dy^2 + dz^2\right] \qquad (8.11)$$

La función a(t) se denomina factor de escala. Si pensamos en términos físicos en lo que está sucediendo, vemos que el tiempo pasa según su ritmo completamente normal. Lo sabemos porque el elemento g00 de la métrica, el coeficiente de dt^2, es −1, al igual que en el espacio de Minkowski. Pero las partes espaciales están multiplicadas por el factor de escala. Si dos objetos, dos galaxias, por ejemplo, están definidos en unas coordenadas espaciales fijas, la distancia entre ellos aumentará a medida que a(t) se incremente. Así pues, tenemos una descripción matemática del espacio en expansión.



Cuando pensamos en torno a la relatividad general, con frecuencia procedemos del siguiente modo: decidimos que queremos crear el modelo de una situación física a partir de unas propiedades dadas. Podemos imaginar, por ejemplo, que tenemos un universo ocupado de un modo uniforme, de acuerdo con una distribución regular de la materia, en lugar de presentar una configuración desigual, no homogénea. Dado que la materia (que curva el espacio-tiempo) se distribuye uniformemente a través del espacio, podemos inferir que la métrica no dependerá de las ubicaciones espaciales xi, sino únicamente de t. Es más, digamos que el espacio (no el espacio-tiempo) presenta una geometría euclidiana plana, en lugar de estar positiva o negativamente curvado. Estos requisitos nos conducen en esencia únicamente a una métrica con la forma de (8.10). Pero no nos dicen cómo se comporta el factor de escala a(t). Eso dependerá de cuánta materia y otras cosas haya en el universo. Está claro que necesitamos una ecuación que relacione la métrica con la materia y la energía. Eso es exactamente con lo que dio Einstein.

EL TENSOR DE ENERGÍA-MOMENTO

En matemáticas, formulamos axiomas y hacemos derivar rigurosamente teoremas a partir de ellos. En ciencia, en cambio, básicamente conjeturamos leyes de la física —o las planteamos como hipótesis, por decirlo de un modo más erudito—, y luego las sometemos a prueba, tanto en cuanto a su consistencia interna, como en cuanto a su validez al contrastarlas con los

datos experimentales. La tarea de Einstein fue la de conjeturar la ecuación que rige el comportamiento de la métrica del espacio-tiempo.

En la gravedad newtoniana, la fuerza es proporcional a la masa del objeto que origina la atracción gravitatoria. Queremos una generalización relativista de esta idea. En la relatividad, la «masa» es simplemente una versión de la energía (es la energía inherente en un objeto cuando está en reposo). Y la energía se unifica con el momento, es el componente «ceroavo» del vector cuadrimomento, como dijimos en el Capítulo 6.

Pero en ese capítulo nos centrábamos en la energía y el momento de un simple objeto que podíamos abstraer convirtiéndolo en un punto. No siempre vamos a ser tan afortunados. A veces nos interesará considerar fuentes de materia y energía diseminadas por el universo, como el interior de una estrella o la distribución de la materia oscura en una galaxia. En la relatividad, tal distribución extendida de materia se llama fluido. No es un término muy feliz, puesto que a menudo consideramos ejemplos que no tienen en absoluto el aspecto de un fluido. El interior de un planeta recibe la denominación de «fluido», aunque sea perfectamente sólido. Una agrupación de fotones moviéndose en direcciones aleatorias es un fluido. Para un estudioso de la relatividad general, «fluido» es cualquier forma de materia extendida por el espacio, por contraposición a la concentrada en un punto.

Para un fluido, más que pensar en masa o energía, podemos pensar en la densidad de energía, expresada habitualmente como ρ (la letra griega rho), y que representa la cantidad de energía por centímetro cúbico (o cualesquiera unidades que prefiera usar). La densidad de energía puede variar a través del espacio y del tiempo, así que generalmente será una función de x_μ. Para un objeto localizado como una estrella o un planeta, la energía total será la integral de la densidad de energía por el espacio.

Pero los fluidos se caracterizan por algo más que la densidad de la energía en cada punto. Puede estar presente también la presión, que puede conceptualizarse como la cantidad de fuerza ejercida por el fluido contra las paredes de un contenedor. (Aun si no está literalmente encerrado en un contenedor, podemos concebir la fuerza que ejercería si lo estuviera). Un fluido puede moverse de manera bastante complicada —pensemos en las corrientes de aire, o de agua—, así que por lo general habrá un vector de

velocidad asociado con el fluido en cada punto. Y puede haber otros tipos de tensiones y esfuerzos si el objeto sufre torsiones o deformaciones con respecto a su forma en estado de equilibrio. Como en el caso de la densidad de energía, todas estas cantidades dependerán en general de cuál sea el punto del espacio-tiempo que estemos considerando.

En la relatividad, todas estas ideas se agrupan en torno al tensor de energía-momento (también conocido como tensor de energía-impulso), expresado habitualmente como Tμν. El tensor de energía-momento sintetiza cuanto queremos saber acerca de la masa, la energía, el momento, la presión, la tensión y otras características relacionadas con la energía en un cúmulo de materia (o de radiación, o de cualquier otra cosa).

Por lo general, como puede imaginar, una expresión para Tμν en forma de cantidades que nos resultaran más familiares tendría un aspecto tremendamente complicado. Pero podemos captar una intuición considerando el caso sencillo de un fluido perfecto: un fluido que tiene el mismo aspecto en todas direcciones en su marco de reposo. En tal caso, las únicas cantidades que necesitamos para especificar el tensor de energía-momento de dicho fluido son la densidad de energía y la presión p. En un espacio-tiempo plano, en el marco de reposo del fluido perfecto, el tensor de energía-momento tendría este aspecto:

$$T_{\mu\nu} = \begin{pmatrix} \rho & & & \\ & p & & \\ & & p & \\ & & & p \end{pmatrix} \qquad (8.12)$$

El elemento 00 es la densidad de energía, y los elementos espaciales en diagonal son iguales a la presión. Para fluidos no perfectos o fluidos en un marco de no reposo, es la locura; la presión puede ser diferente en cada diferente dirección, y pueden darse términos fuera de la diagonal, resultantes de la tensión y el esfuerzo. Pero ya bastante estamos poniendo a prueba nuestro cerebro, así que será suficiente si nos ceñimos a la forma simple

(8.12). Tanto ρ como p pueden ser funciones de xμ, por lo que esta expresión puede encerrar mucha información. Los fluidos perfectos pueden describir planetas, estrellas, o la materia oscura y la energía oscura que llena el espacio.

LA ECUACIÓN DE EINSTEIN

Por tanto, para generalizar la gravedad newtoniana al contexto relativista, queremos formular una ecuación que relacione la métrica del espacio-tiempo con el tensor de energía-momento. Se trata de una extensión de la unificación que vimos cuando relacionamos la energía de una partícula con su momento. En la relatividad general, la gravedad no la crea solo la masa, la crean todo tipo de magnitudes diferentes como la energía, la presión, la tensión, etcétera.

¿Cómo lo hacemos, entonces? Tanto gμν como Tμν son tensores con dos subíndices, y además ambos son simétricos (gμν = gνμ y Tμν = Tνμ). Así que, como primera aproximación, imaginemos que son proporcionales entre sí:

$$g_{\mu\nu} = \alpha T_{\mu\nu}, \qquad (8.13)$$

donde α es una constante de proporcionalidad. Siempre que tengamos una ecuación que implique tensores, ambas partes deben tener los mismos índices libres en cada término, de lo contrario no estaríamos haciendo entrar en la ecuación la misma forma de tensor.

Puede parecer una conjetura bastante tonta, pero estamos intentando darle a conocer el tipo de cosas que pasarían por la mente de un físico teórico, aunque sea por un milisegundo, antes de ser descartadas. Inmediatamente vemos que no puede ser correcta, porque en el espacio vacío tendríamos Tμν = 0 (forma abreviada de asignar 0 en cada posición de elemento), y desde luego no queremos tener gμν = 0. Queremos que nuestra ecuación nos

proporcione la métrica de Minkowski en el espacio vacío, o más específicamente en situaciones en que no hubiera gravedad.

Usemos un poco más el cerebro. Aunque (8.13) parece matemáticamente legítima —relaciona en una ecuación dos tensores simétricos con dos subíndices—, no tiene sentido físico. Intuitivamente, esa ecuación implicaría que la energía-momento está de algún modo creando la métrica. Eso no es lo que queremos, en absoluto. Queremos que la energía-momento desvíe la métrica: no que cree el espacio-tiempo, sino que curve el espacio-tiempo. Si no hay fuentes para ($T\mu\nu = 0$), el espacio-tiempo puede ser plano, pero si incluimos ahí un planeta o una estrella, el espacio-tiempo puede desviarse.[1]

Si lo pensamos, lo que realmente queremos es que el tensor de energía-momento actúe como una fuente para derivadas de la métrica, no de la métrica misma. Derivadas de valor diferente a cero son la forma en que caracterizamos la desviación. En el Capítulo 4 mencionamos Laplace introdujo el campo potencial gravitatorio como una forma de concebir la gravedad newtoniana. En ese contexto, la fuerza gravitatoria depende de la derivada del potencial, no del potencial mismo. En nuestro nuevo contexto relativista, deberíamos concebir el tensor métrico aproximadamente análogo al potencial gravitatorio, cuyas derivadas nos proporcionan la fuerza, más que la fuerza gravitatoria misma.

De modo que lo que buscamos es una cantidad que sea un tensor simétrico con dos subíndices (para poder establecerlo como proporcional a $T\mu\nu$), que podamos extender a la métrica y sus derivadas.

$$\left(\begin{array}{c}\text{tensor simétrico}\\ \text{construido a partir}\\ \text{de la métrica y}\\ \text{sus derivadas}\end{array}\right)_{\mu\nu} = \alpha T_{\mu\nu} \qquad (8.14)$$

Ya casi tenemos algo así en nuestra posesión. El tensor de curvatura de Riemann $R\lambda\sigma\mu\nu$ (para el que ahora utilizamos letras griegas al estar en el

espacio-tiempo) es un tensor elaborado a partir de derivadas de la métrica. El único problema es que tiene demasiados índices. Pero hay otro tensor, el tensor de Ricci, que podemos formular sumando simplemente el primer y tercer índice del tensor de Riemann. El tensor de Ricci recibe su nombre del matemático italiano Gregorio Ricci-Curbastro, quien ideó también los fundamentos del cálculo de tensores y gran parte del aparato de la geometría riemanniana moderna. En un influyente artículo para una revista científica, un texto de 1900 escrito con la colaboración de su antiguo alumno Tullio Levi-Civita y que Einstein utilizaría para el estudio de los tensores, por algún motivo desconocido firmó como «G. Ricci», sin el compuesto «-Curbastro». Todos sus demás artículos incluían su apellido completo, esta fue la única excepción. Quizá sospechara que el tensor que iba a presentar merecería un nombre breve y susceptible de recordarse.

Utilizando el convenio de suma de Einstein, el tensor de Ricci se formula:

$$R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu} = R^{0}{}_{\mu 0\nu} + R^{1}{}_{\mu 1\nu} + R^{2}{}_{\mu 2\nu} + R^{3}{}_{\mu 3\nu} \qquad (8.15)$$

Hemos trastocado un poco nuestras letras griegas, pero así está bien; al fin y al cabo, no son más que etiquetas arbitrarias. Con tal de que nuestra elección de índices sea coherente (en una misma agrupación de tres) en el seno de la ecuación, todo irá bien. El tensor de Ricci también es simétrico, $R\mu\nu = R\nu\mu$.

Desde luego parece que estamos muy cerca del objetivo. Obviamente, deberíamos formular una ecuación de la forma:

$$R_{\mu\nu} = \alpha T_{\mu\nu} \qquad (8.16)$$

donde α es, como antes, una constante de proporcionalidad. Esta hipótesis es mucho más razonable que (8.13). Sigue la forma general (8.14), estableciendo un tensor simétrico de dos índices construido a partir de la métrica y sus derivadas, igual al tensor de energía-momento. Y en el espacio vacío, cuando $T\mu\nu = 0$, predice que $R\mu\nu = 0$, lo cual es ciertamente coherente con el espacio plano de Minkowski (en el que todos los elementos del tensor

de Riemann desaparecen, por lo que el tensor de Ricci a buen seguro también).

De hecho, es tan razonable, que el propio Einstein propuso esta ecuación como posible fundamentación para la relatividad general en octubre de 1915. Y casi funciona, pero no del todo.

El problema surge cuando pensamos en algo que sabemos acerca de la energía: que se conserva. La cuestión de la conservación de la energía en la relatividad general es delicada, porque la energía puede transferirse en ambos sentidos entre la materia y la curvatura del espacio-tiempo. Pero una vez tenemos esto en cuenta, se da una fuerte restricción acerca del modo en que el tensor de energía-momento puede cambiar con el tiempo. Y esta restricción no la cumple el tensor de Ricci. Así, si (8.16) fuera la ecuación correcta, o bien la energía no se conserva en absoluto, o casi ninguna métrica espacio-temporal resolvería en realidad la ecuación.

La buena noticia es que la cuestión tiene fácil arreglo. La mala es que los detalles requieren profundizar un poco más en el meollo de los tensores y la curvatura, cosa que hemos dejado para el Apéndice B. El truco fundamental es que podemos definir una métrica inversa, $g\mu\nu$, que está relacionada con la métrica, pero con superíndices en lugar de subíndices. (Si sabe de matrices, esto es literalmente la matriz inversa de la métrica). Utilizando la métrica inversa podemos definir una función de espacio-tiempo, el escalar de curvatura de Ricci, mediante:

$$R = g^{\mu\nu} R_{\mu\nu} \qquad (8.17)$$

Aquí no hay índices libres, ya que hemos sumado tanto μ como ν en la parte de la derecha. Pero podemos multiplicarlo por la métrica $g\mu\nu$ para obtener un tensor simétrico diferente de dos índices, construido a partir de la métrica y sus derivadas. Entonces, si fuera usted Einstein garabateando febrilmente en noviembre de 1915, podría intentar hallar una combinación de $R\mu\nu$ y $Rg\mu\nu$ que tuviera las propiedades adecuadas que la hicieran proporcional a $T\mu\nu$ sin vulnerar el principio de conservación de la energía. Hay una única respuesta correcta, que en la actualidad conocemos como ecuación de Einstein:

$$_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu} \qquad (8.18)$$

La combinación de la parte izquierda es el tensor de Einstein. Podríamos crear un nuevo símbolo para representarlo, pero tal como está ya resulta una expresión suficientemente sencilla, como combinación del tensor de Ricci y el escalar de curvatura. Esta es la forma final de la ecuación de campo para la relatividad general, presentada por Einstein en una disertación ante la Academia de Ciencias de Prusia, el 25 de noviembre de 1915.[2]

El físico John Wheeler sintetizó la relatividad del siguiente modo: «el espacio-tiempo le dice a la materia cómo moverse; la materia le dice al espacio-tiempo cómo curvarse». La primera parte de esta máxima se concretiza en la idea de que las partículas libres se desplazan siguiendo líneas geodésicas, y las partículas no libres (aquellas que están sujetas a una fuerza diferente a la gravedad) se desvían del movimiento geodésico del mismo modo, en gran medida, en que las partículas no libres se apartan al acelerarse de las líneas rectas en la mecánica newtoniana. La segunda parte se manifiesta en la ecuación de Einstein, que podemos resolver para que nos diga cuál será la métrica del espacio-tiempo en cualquier situación que nos interese. Esta ecuación ha terminado por predecir correctamente la evolución del universo, la existencia de agujeros negros, la propagación de las ondas gravitatorias y otros fenómenos de los que Einstein no tenía indicio alguno en su época. Tal es el poder auténtico de una buena teoría científica: encierra mucho más conocimiento del que pueden sospechar las personas que la formulan.

La forma en que hemos expresado la ecuación de Einstein no incluye una constante indeterminada de proporcionalidad, sino un factor específico, 8πG, donde G es la misma constante que aparece en la ley de gravitación de Newton. No podemos deducir este valor por puro raciocinio, o ni siquiera buscando su coherencia con otros preciados principios como la conservación de la energía, sino que tenemos que recurrir a los datos experimentales. Lo que hizo Einstein fue considerar el «límite de campos débiles», en que la gravedad es débil y el espacio-tiempo es casi plano, pero no por completo. En tales circunstancias, sabemos que una buena teoría de la gravedad debería

reproducir la ley de la inversa del cuadrado de Newton, y para que esto se cumpla, la constante en la ecuación de Einstein (8.18) tiene que ser $8\pi G$. Lo asombroso es que esta ecuación, con su constante numérica establecida al medir la caída de las manzanas de los árboles y el movimiento de los planetas en el sistema solar, realiza predicciones sobre lo sucedido en los primeros minutos posteriores al Big Bang, y lo hace con gran éxito.

PRINCIPIO DE ACCIÓN

En los Capítulos 3 y 4 vimos que hay formas de formular la física clásica que tienen una apariencia diferente, pero que son matemáticamente equivalentes: la mecánica newtoniana, la mecánica lagrangiana y la mecánica hamiltoniana. La relatividad general es una teoría clásica, de modo que no debería sorprendernos que pueda expresarse también de múltiples formas equivalentes. Observemos el camino lagrangiano, siguiendo el principio de mínima acción. Resulta ser un modo especialmente adecuado para pensar en la teoría de la relatividad, siendo idóneo como es para tratar el espacio y el tiempo en pie de igualdad.

En nuestra primera aproximación al principio de acción, partimos de una partícula determinada por una posición x y una velocidad $v = dx/dt$. Definíamos un lagrangiano L como una función de x y v, en particular como la energía cinética menos la energía potencial. La acción es la integral del lagrangiano en el tiempo:

$$S = \int L\left(x, \frac{dx}{dt}\right) dt \qquad (8.19)$$

La trayectoria efectiva tomada por una partícula real es aquella que minimiza esta acción en comparación con otras trayectorias entre los mismos puntos de partida y de llegada.

Ahora tenemos una situación algo diferente. Más que una partícula con una ubicación en el espacio, nos interesa la dinámica del tensor métrico. La relatividad general es un ejemplo de teoría de campos, ya que el tensor métrico gμν(t, xi) es un campo con un valor en cada punto del espacio-tiempo, más que una partícula con una ubicación en algún lugar. En una teoría de campos, definfimos el lagrangiano elaborando una función llamada densidad de Lagrange , e integrándola en todo el espacio:

$$L(t)=\int \mathcal{L}\left(t,\, x^{i}\right) d^{3}x \qquad (8.20)$$

La notación $d^3x = dx^1dx^2dx^3$ indica que estamos integrando en las tres dimensiones del espacio. Cuando integramos una función espacio-temporal (la densidad de Lagrange) en el espacio, tenemos al final una función de tiempo (el lagrangiano mismo). La acción es la integral de en el tiempo, que es justamente la integral de en el espacio-tiempo:

$$S=\int L\,dt=\int \mathcal{L}\,d^{4}x \qquad (8.21)$$

Hemos hallado ya la ecuación de Einstein, pero imaginemos que aún no lo hubiéramos hecho y estuviéramos intentando formularla utilizando el principio de mínima acción. Nuestra tarea está clara: necesitamos estimar la densidad de Lagrange apropiada. Debería poder elaborarse a partir de la métrica y sus derivadas, al igual que la densidad de Lagrange para una partícula se elabora a partir de la posición y sus derivadas (en particular, la velocidad). Pero la buena noticia es que estamos intentando estimar una función escalar —un tensor con cero índices, si lo prefiere—, en lugar de un tensor de dos índices, como miembro de la parte izquierda de (8.14). Hay menos maneras de tener cero índices en un tensor que dos índices, lo cual nos facilita considerablemente las cosas.

De hecho, esencialmente, solo hay una posibilidad: el escalar de curvatura de Ricci R. Es la estimación obvia en el caso de la densidad de Lagrange para la métrica, gravedad = R, básicamente porque no hay más opción. Deberíamos

incluir también una densidad de Lagrange para la materia, pero no necesitamos explicitarla, dependerá del tipo de materia que nos interese. Y necesitamos incluir la constante de Newton G en algún lugar para que la fuerza de gravedad funcione correctamente. La respuesta final resulta:

$$S = \int \left(\frac{1}{16\pi G} R + \mathcal{L}_{\text{materia}} \right) \widehat{d^4 x} \qquad (8.22)$$

¡Eso es! Esta es la acción que, cuando buscamos métricas espacio-temporales que la minimicen, nos dice que obedecerán a la ecuación de Einstein (8.18). Hay un detalle que hemos omitido para mayor simplicidad, esto es, que el «elemento volumen» de la integral sufra una leve alteración en el espacio-tiempo curvo, por lo que lo hemos expresado como d^4x, en lugar simplemente de d^4x, a modo de recordatorio.[3]

La belleza de la formulación de la acción debería resultar evidente: era bastante más fácil estimar la correcta densidad escalar de Lagrange que el tensor correcto para la ecuación de Einstein, y los preciados principios como el de conservación de la energía surgen automáticamente, en lugar de permanecer como cosas objeto de preocupación y pendientes de verificación. Por supuesto, hay que ser bastante ágil con las matemáticas para intuir, ya de entrada, que hay que pensar en el principio de acción, y luego abrirse paso manejándose con las fórmulas (lo cual hemos evitado prudentemente aquí) para hacer derivar realmente la ecuación de Einstein a partir de la acción.

Einstein era ciertamente ágil con las matemáticas, pero no tanto como su colega David Hilbert, que fue uno de los grandes matemáticos de comienzos del siglo XX. (El «espacio de Hilbert» aparece como un concepto crucial en la mecánica cuántica). En el verano de 1915, cuando Einstein se acercaba ya a la forma final de la relatividad general, pero no la había alcanzado aún del todo, Hilbert lo invitó a impartir una serie de clases magistrales en la Universidad de Gotinga. Ambos tuvieron ocasión de hablar largo y tendido acerca del espacio-tiempo curvo, Einstein incluso se alojó en casa de Hilbert. Tras el regreso de Einstein a Berlín, mantuvieron una extensa correspondencia. Al final, ambos obtuvieron la ecuación (8.18) de forma casi

simultánea, Einstein mediante el método de «conjetura y comprobación» y Hilbert a través de la elegante formulación de la acción.

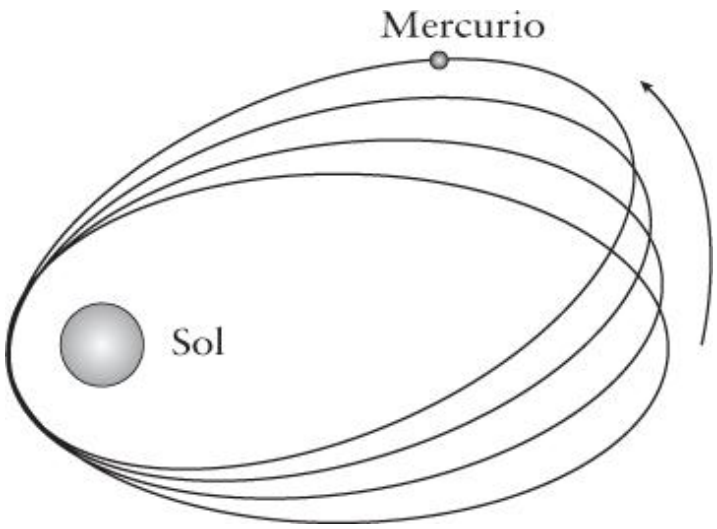
No es más que una cuestión de discusión académica dilucidar si Hilbert realmente obtuvo la ecuación de campo completa unos días antes de que lo hiciera Einstein, o el grado de utilidad que supuso para Einstein la información recibida a través de la correspondencia con Hilbert. Faltan cartas, algunos artículos fueron actualizándose durante el proceso de edición, y hay que contar también con el habitual amasijo de minucias que forman parte de la historia. Lo que está claro es que fue Einstein quien tuvo la idea básica de abordar la gravedad en términos de curvatura del espacio-tiempo, y quien presentó primero la ecuación final en público, sobre la base de criterios físicos razonables. Por eso los físicos se refieren a (8.18) como la «ecuación de Einstein», y a (8.22) como la «acción de Einstein-Hilbert». Se trata de una correspondencia relativamente exacta entre las etiquetas asignadas y la atribución histórica merecida, que los científicos no siempre obtienen.

CONSECUENCIAS EMPÍRICAS

A diferencia de la mayor parte de teorías físicas, el desarrollo de la relatividad general no fue impulsado principalmente por la necesidad de explicar algún tipo de anomalía experimental desconcertante, sino por la búsqueda de consistencia teórica. Einstein tenía en mente algunas características acerca del funcionamiento de la gravedad, como la ley de la inversa del cuadrado o el principio de equivalencia, y estaba muy familiarizado con la estructura básica de la relatividad. Era cuestión de conciliar estas exigencias teóricas entre sí, cosa que finalmente consiguió postulando que la gravedad era una manifestación de la curvatura del espacio-tiempo.

Pero una vez hecho esto, y obtenida la ecuación de campo, llegaba el momento de volver al terreno de lo empírico: plantear predicciones experimentales, levantarse de la mesa y probarlas.

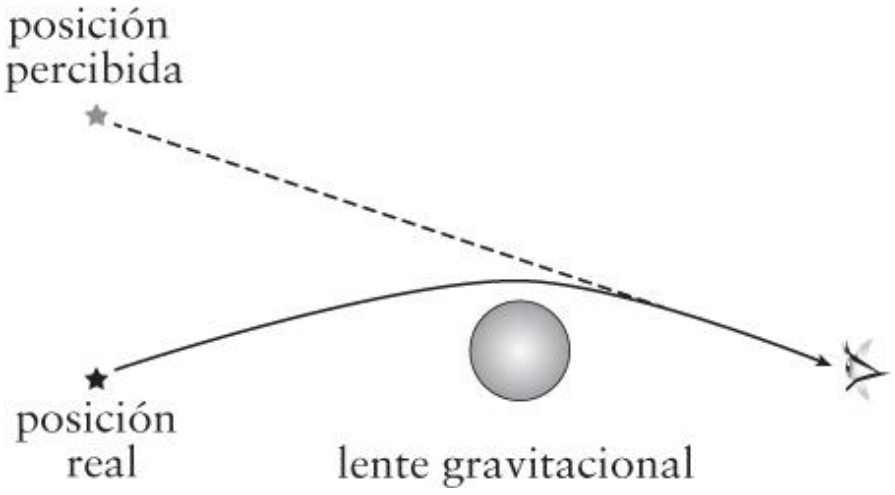
Un campo de pruebas fue la precesión de la órbita de Mercurio, lo cual representaba una pequeña trampa, por cuanto la situación observacional era ya suficientemente conocida. Kepler había postulado que los planetas se desplazan describiendo elipses perfectas, y Newton había propuesto una teoría según la cual tal cosa sucedía de forma natural si hay un único planeta que se mueve sin sufrir perturbación alrededor de un sol perfectamente esférico. En el universo real, se dan leves tirones originados por los campos gravitatorios de los demás planetas, que hacen que las órbitas experimenten cierta precesión, lo que hace que se desplace lentamente la dirección hacia la que se orienta el eje de la elipse. Traducido en números, la gravedad newtoniana predice que la órbita de Mercurio se desplaza 0,148 grados cada siglo.



Hacia finales del siglo XIX, los astrónomos habían medido la precesión de Mercurio en 0,160 grados por siglo. Es una discrepancia de 0,012 grados/siglo, pequeña, pero lo bastante importante como para que no pudiera deberse a un error aleatorio. Urbain Le Verrier, astrónomo francés que había propuesto la existencia del planeta Neptuno para explicar una anomalía análoga con respecto a órbita de Urano, intentó repetir éxito sugiriendo que podría haber un nuevo planeta que describiera una órbita interna con respecto a la de Mercurio. El hipotético planeta recibió incluso nombre: Vulcano. Diversos equipos de investigadores afirmaron haberlo descubierto, pero todo ello se reveló erróneo en exámenes más detenidos.

Einstein se dio cuenta de que la relatividad general realizaba una buena labor a la hora de recuperar las predicciones de la gravedad newtoniana, pero la coincidencia no era exacta. Había pequeñas correcciones que cada vez se revelaban más importantes para campos gravitatorios más fuertes. Cabía esperar por tanto que, de entre todos los planetas, las más destacables fueran las de Mercurio, teniendo en cuenta que es el más cercano al sol. Einstein se propuso calcular la discrepancia en su precesión predicha por su teoría, y obtuvo como resultado 0,012 grados por siglo, que coincidía con exactitud con los datos conocidos. Apenas podemos imaginar el júbilo de Einstein (y su alivio), después de años de enconado esfuerzo con el análisis de tensores y otras abstracciones matemáticas, cuando comprobó que su teoría proporcionaba una explicación perfecta a una anomalía observacional persistente.

Es estupendo resolver un enigma de larga duración, pero en los círculos científicos se considera más impresionante ser capaz de predecir un fenómeno que aún no se ha observado, y salir luego a verificarlo. Tal fue el caso de la deflexión de la luz, o lente gravitacional, que Einstein predijo aun antes de obtener la ecuación de campo completa para la relatividad general. Ello fue posible porque realmente es consecuencia del principio de equivalencia. Desde la perspectiva de un cohete espacial acelerado, un rayo de luz parecería doblarse, solo por el movimiento cambiante de la nave. Y si esto es cierto para un cohete que experimenta aceleración, también debería serlo para un observador inmóvil sobre la superficie de un planeta dotado de campo gravitatorio.



O, por mejor decir aún, si la luz pasa cerca de la superficie de un objeto con elevada gravedad como el sol. El único problema estriba en que el sol es muy brillante, y es difícil ver estrellas en sus proximidades, cosa que tendríamos que hacer con el fin de medir sus posiciones y determinar si su luz ha sufrido deflexión. El misterio podía desvelarse esperando a un eclipse total de sol, como el que oportunamente tuvo lugar en 1919. Una expedición organizada por el astrofísico británico Arthur Eddington tomó fotografías de las posiciones estelares cerca del sol durante el eclipse y verificó la predicción de Einstein referente a la desviación de la luz.

Fue este resultado empírico, más que su propuesta teórica original, lo que convirtió a Einstein en una celebridad internacional. Los resultados de las observaciones de Eddington ocuparon las primeras páginas de los periódicos de todo el mundo, incluido The New York Times. En su elaborado titular múltiple, el rotativo proclamaba acerca de la relatividad general, entre otras cosas, que «sir Oliver Lodge dice que triunfará, y que a los matemáticos les espera una época terrible». Se equivocaban, los matemáticos estaban encantados.

Al igual que los físicos y los astrónomos. En la actualidad, las observaciones de las lentes gravitacionales han madurado hasta alcanzar un nivel científico de alta precisión y convertirse en un instrumento fundamental de la caja de herramientas del cosmólogo moderno. Podemos mirar a las profundidades del universo y observar las posiciones de un gran número de galaxias, sirviéndonos de los patrones estadísticos de las lentes gravitacionales para inferir la presencia de concentraciones de materia, en gran parte «materia oscura». La relatividad general muestra sin ambigüedades que todas las formas de energía provocan la curvatura del espacio-tiempo, lo que hace que la deflexión de la luz sea un método especialmente útil para rastrear la distribución de la materia a través del espacio.

Después de estas pruebas clásicas, realizadas antes o poco después de la formulación de la relatividad general, muchos otros fenómenos fueron descritos por la teoría: la luz, a medida que se aleja de un cuerpo gravitatorio, perderá energía y su longitud de onda se modificará hacia valores más altos, fenómeno que se conoce como corrimiento al rojo; la materia en movimiento

ocasionará ondulaciones en la curvatura del espacio-tiempo que se propagan hacia el exterior a la velocidad de la luz y que se conocen como ondas gravitatorias; densas acumulaciones de materia colapsarán por efecto de su atracción gravitatoria, creando regiones del espacio de las que ni siquiera la propia luz puede escapar, y que son conocidas como agujeros negros; y el universo mismo no está quieto, ya que el espacio, ocupado por materia, debe expandirse o contraerse, y en la década de los años 1920 Edwin Hubble estableció la expansión del universo. Todas estas son implicaciones impresionantes de la relatividad general que ahora se han observado ya con primorosa precisión.

Es una lástima que Albert Einstein, que murió en 1955, no viviera para presenciar el notable apogeo en la aplicación de la relatividad general a toda una serie de importantes problemas de la física. Durante su vida, las observaciones astrofísicas no estuvieron por lo general a la altura de la tarea investigadora en torno a las situaciones en las que la relatividad entra en juego. Einstein no recibió el Premio Nobel por la relatividad, ni tampoco Hubble por la expansión del universo, debido en parte a un prejuicio contra las cuestiones astronómicas.

Los tiempos han cambiado. Presentamos una lista de Premios Nobel que han sido concedidos por descubrimientos o fenómenos en los cuales la relatividad general desempeña un papel central, hasta 2021:

•1978: Descubrimiento de la radiación cósmica de fondo.

•1993: Púlsar binario, prueba indirecta de las ondas gravitatorias.

•2006: Fluctuaciones y espectro de fondo cósmico de microondas.

•2011: Aceleración de la expansión del universo.

•2017: Observación directa de las ondas gravitatorias.

•2019: Evolución de las galaxias y del universo.

•2020: Teoría y observaciones en torno a los agujeros negros.

Como podemos ver, el ritmo se acelera. Considerada durante mucho tiempo como un gran logro intelectual, pero no como una preocupación fundamental para el profesional de la física, la relatividad general ocupa cada vez más un lugar central entre los apasionantes temas de la investigación contemporánea. Nos gusta imaginar que Einstein estaría encantado.

1 Aun sin fuentes, el espacio-tiempo puede no ser plano. Podría haber ondas gravitacionales atravesando el espacio vacío, por ejemplo. Pero si el espacio-tiempo es plano, no puede haber ninguna fuente.

2 Hay otra cosa que usted podría hacer: sumar un término proporcional a la métrica misma, $\Lambda g_{\mu\nu}$, donde Λ es una constante. Einstein exploró esta posibilidad en 1917, llamando a Λ la constante cosmológica . Las pruebas de una constante cosmológica de valor diferente a cero se reunieron finalmente en 1998, cuando los astrónomos descubrieron la aceleración del universo. Pero el valor de Λ obtenido en las mediciones es tan pequeño, que normalmente podemos ignorarlo, a menos que estemos dedicándonos a la cosmología.

3 Si necesita saberlo, el elemento volumen correcto es $d^4x = \sqrt{-g}\, d^4x$, donde g es el determinante de matriz del tensor métrico.

NUEVE

AGUJEROS NEGROS

La ecuación de Einstein para la relatividad general encierra un tesoro de información en un compacto embalaje. Podemos dar las gracias al milagro de la formulación ingeniosa. La ecuación está pensada para determinar la métrica del espacio-tiempo, $g\mu\nu(x)$, pero está expresada en la forma del tensor de Ricci, elaborado a partir del tensor de curvatura de Riemann. Estos tensores se definen en términos de la métrica, ciertamente, pero si tuviéramos que formular explícitamente esta dependencia en toda su profusa gloria, nos veríamos llevados a un aluvión de términos que llenaría una página entera de símbolos matemáticos.

El propio Einstein se sintió lo bastante impresionado, o intimidado, por la complejidad de su propia ecuación como para recurrir sin dudarlo a métodos de aproximación, como el límite newtoniano, con el fin de deducir predicciones experimentales. La ecuación parecía demasiado complicada para resolverla con total exactitud, aun en situaciones simplificadas.

Todo ello no disuadió a Karl Schwarzschild. Físico y astrónomo consumado, Schwarzschild servía en 1915 en el ejército alemán durante la Primera Guerra Mundial. En los frentes francés y ruso pasaba el tiempo calculando las trayectorias de los obuses, pero con ocasión de un permiso tuvo la oportunidad de asistir a una de las disertaciones de Einstein en la Academia de Ciencias de Prusia, y quedó fascinado por la relatividad general. De retorno a sus deberes militares, a finales de diciembre de 1915, Schwarzschild escribió a Einstein una carta con la primera solución exacta a la ecuación de Einstein, y en la que describía la métrica en la proximidad de un planeta o estrella esféricos. Por desgracia, Schwarzschild contrajo en el frente una rara enfermedad de la piel, de la que murió menos de seis meses después, a la edad de cuarenta y dos años. Los físicos tardaron décadas en

aceptar una implicación de su descubrimiento que era completamente imprevista y se antojaba pasmosa: la predicción de agujeros negros en la relatividad general.

LA SOLUCIÓN DE SCHWARZSCHILD

Schwarzschild buscaba el equivalente a la ley de la inversa del cuadrado de Newton en el sistema solar. En la relatividad general, esto significaba encontrar una solución a la ecuación de Einstein para la métrica en el espacio vacío alrededor de un cuerpo esférico aislado como el sol. Calculando líneas geodésicas de esa métrica, conocemos cosas acerca de las órbitas planetarias, la deflexión de la luz y otras predicciones de la relatividad general. Sería suficientemente estupendo presentar la solución de Schwarzschild tal cual, en todo su esplendor, para a continuación hablar un poco de algunas de sus implicaciones. Pero es más divertido rehacer el camino hasta ella, resiguiendo sus pasos básicos, para mostrar cómo aborda un físico teórico un problema como este.

Sabemos que la métrica del espacio-tiempo plano de Minkowski, expresada en coordenadas cartesianas (t, x, y, z), adopta la forma:

$$g_{\mu\nu} = \begin{pmatrix} g_{tt} & & & \\ & g_{xx} & & \\ & & g_{yy} & \\ & & & g_{zz} \end{pmatrix} = \begin{pmatrix} -1 & & & \\ & +1 & & \\ & & +1 & \\ & & & +1 \end{pmatrix} \quad (9.1)$$

Como hicimos anteriormente, hemos omitido el 0 en los elementos que están fuera de la diagonal, pero hay que recordar que están ahí.

Una métrica curva tendrá algunos (o quizá todos) de los elementos métricos $g\mu\nu$ como funciones de algunas (o quizá de todas) de las coordenadas. Esto suena a que las cosas van a complicarse mucho muy deprisa. Pero nos ayuda enormemente el hecho de que la situación física por la que nos cuestionamos (el espacio-tiempo en la proximidad de un cuerpo esférico) tiene que ser también esféricamente simétrica. En la práctica, esto significa que deberíamos esperar que la métrica dependa de la distancia con respecto al origen, $r = \sqrt{x^2 + y^2 + z^2}$, más que de los valores individuales de x, y o z.

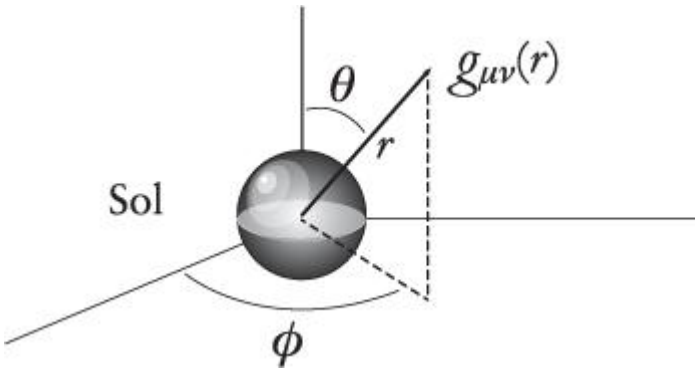
Esto nos sugiere un primer paso: transformar la parte espacial de nuestras coordenadas, de cartesiana (x, y, z) a esférica (r,θ,ϕ). En la ecuación (7.12) ya formulamos la métrica para un espacio plano euclidiano en coordenadas esféricas. A partir de ella, resulta más sencillo ver el aspecto que adopta la métrica del espacio-tiempo de Minkowski para coordenadas (t,r,θ, ϕ):

$$g_{\mu\nu} = \begin{pmatrix} g_{tt} & & & \\ & g_{rr} & & \\ & & g_{\theta\theta} & \\ & & & g_{\phi\phi} \end{pmatrix} = \begin{pmatrix} -1 & & & \\ & +1 & & \\ & & r^2 & \\ & & & r^2(\text{sen}\theta)^2 \end{pmatrix} \quad (9.2)$$

Simplemente, hemos tomado la parte espacial de la métrica y la hemos reemplazado por la versión con coordenadas esféricas. Subrayemos que seguimos con la métrica del espacio-tiempo plano de Minkowski, nada que ver todavía con la gravedad. Tan solo estamos expresándola en coordenadas que nos ayuden a trazar el camino hacia la configuración que debería tener la métrica en la proximidad de un cuerpo gravitatorio.

En coordenadas esféricas, todo lo que necesitamos expresar es el modo en que la métrica depende de la coordenada radial r; la dependencia con respecto a las coordenadas angulares (θ,ϕ) viene determinada por la simetría esférica. Es decir, la métrica no dependerá en absoluto de ϕ, y solo dependerá de θ a través de un factor $(\text{sen}\theta)^2$, en el elemento $g\phi\phi$ de la esquina inferior derecha. Y hay más simplificaciones. Una de ellas es que

estamos buscando soluciones estáticas: el espacio-tiempo está ahí, quieto, sin evolucionar en el tiempo. De modo que los elementos de la métrica tampoco dependerán de t. Podríamos también contemplar la posibilidad de que los términos que no están en la diagonal tuvieran valor diferente a cero, pero es más fácil imaginar al principio algo que sea simple, y en este caso eso nos sirve.



Quizá pueda sentirse incomodado por tanta conjetura. Suena poco científico. Pero cuando se trata de resolver ecuaciones, conjeturar qué características debe tener la solución es perfectamente legítimo. No estamos intentando dilucidar cada una de las soluciones particulares a la ecuación de Einstein, sino tan solo un caso concreto. Al final del proceso, si nuestras conjeturas nos conducen a una métrica por la que podemos calcular el tensor de Riemann y el tensor de Ricci, y comprobar que todo ello cumple con la ecuación de Einstein, ninguna de las triquiñuelas de que nos hayamos valido para llegar a ello tendrá importancia alguna.

Finalmente, pues, observemos con detenimiento (9.2) y pensemos en lo que los factores de r^2 están diciéndonos realmente. Están ahí para reflejar el hecho de que las distancias físicas que siguen las direcciones angulares ascienden directamente con el valor de r. Tenemos aquí una sutileza propia de la relatividad general: en realidad no estamos eligiendo primero unas coordenadas y resolviendo luego la métrica; estamos haciendo ambas cosas simultáneamente. Las coordenadas no significan nada hasta que la métrica les confiere un significado, y los elementos de la métrica solo tienen sentido en relación con unas coordenadas concretas.

Lo que esto significa es que podemos definir «r^2» simplemente como «la cantidad que aparece en los elementos angulares $g\theta\theta$ y $g\phi\phi$ del tensor métrico». De forma equivalente, definimos r de tal modo que el área de una esfera de r dado sea A = 4 y su circunferencia sea C = 2. Así, no tenemos por qué dejar que estos elementos sean funciones arbitrarias y resolver luego esas funciones. Podemos simplemente afirmar $g\theta\theta = r^2$ y $g\phi\phi = r^2(\text{sen})^2$, como aparece ya en (9.2), tomando eso como la asignación del significado de la coordenada r, siempre y cuando estemos en una situación estática, esféricamente simétrica.[1]

Esto nos deja con una métrica con la siguiente forma:

$$g_{\mu\nu} = \begin{pmatrix} g_{tt} & & & \\ & g_{rr} & & \\ & & g_{\theta\theta} & \\ & & & g_{\phi\phi} \end{pmatrix} = \begin{pmatrix} -A(r) & & & \\ & +B(r) & & \\ & & r^2 & \\ & & & r^2(\text{sen}\theta)^2 \end{pmatrix} \quad (9.3)$$

No está mal. Todas nuestras conjeturas basadas en la perspicaz intuición física nos han dejado con una métrica bastante simple que solo tiene dos funciones indeterminadas de una variable, A(r) y B(r).

Por desgracia, este es el límite al que puede llevarnos la perspicacia. Llegados a este punto, tenemos que apechugar y calcular (es decir, Schwarzschild calculó) el tensor de Riemann primero y el tensor de Ricci después. No vamos a explicitarlo, si bien en el Apéndice B encontrará toda la tecnología necesaria, así que dese una oportunidad. Por ahora aceptaremos que puede hacerse, y a partir de aquí podemos calcular la parte de la izquierda de la ecuación de Einstein. Las expresiones resultantes las formularemos prosiguiendo con A(r) y B(r) y sus derivadas con respecto a r.

Luego tenemos que disponer tales expresiones en igual forma a las de la parte derecha de la ecuación de Einstein, es decir, las que representan el tensor de energía-momento. Pero las buenas noticias siguen saliendo a

nuestro encuentro: por el momento solo nos preocupa cuál sea la métrica fuera del cuerpo gravitatorio, donde el espacio es vacío y Tμν = 0. Dentro del sol o de un planeta pueden suceder cosas más complicadas, pero para el propósito que nos ocupa tales cuerpos son solo fuentes esféricas que provocan la deformación del espacio-tiempo a su alrededor.

Schwarzschild hizo todo esto, y para su sorpresa las ecuaciones resultantes no eran tan difíciles de resolver. (Para ser honrados, tampoco eran tan triviales, pero Schwarzschild era un tipo muy inteligente). Esta es la expresión que encontró para A(r) y B(r):

$$A(r)=\frac{1}{B(r)}=1-\frac{2GM}{r} \qquad (9.4)$$

En otras palabras, la métrica completa de Schwarzschild es:

$$g_{\mu\nu}=\begin{pmatrix} g_{tt} & & & \\ & g_{rr} & & \\ & & g_{\theta\theta} & \\ & & & g_{\phi\phi} \end{pmatrix}=\begin{pmatrix} -\left(1-\frac{2GM}{r}\right) & & & \\ & +\left(1-\frac{2GM}{r}\right)^{-11} & & \\ & & r^2 & \\ & & & r^2\ (\mathrm{sen}\theta)^2 \end{pmatrix} \qquad (9.5)$$

O, en forma de elemento de línea:

$$ds^2=-\left(1-\frac{2GM}{r}\right)dt^2+\frac{1}{\left(1-\frac{2GM}{r}\right)}dr^2+r^2d\theta^2+r^2\left(\mathrm{sen}\theta\right)^2d\phi^2 \qquad (9.6)$$

¿No es precioso? Esta es la solución exacta a la ecuación de Einstein en el espacio vacío con simetría esférica. De hecho, es la única solución. Habíamos descrito la vía de la conjetura, pero un análisis más riguroso puede demostrar que la de Schwarzschild es la única métrica esféricamente simétrica posible que resuelve la ecuación de Einstein en el vacío. Podría objetar a ello que la métrica de Minkowski es otra posible solución, pero no tiene más que establecer M = 0 en (9.5) o (9.6) y comprobar una vez más que Minkowski es un caso especial.

En estas expresiones para la métrica de Schwarzschild, G es por supuesto la constante de gravitación de Newton, y M es la masa del objeto cuyo campo gravitatorio hemos calculado. Pero, si hay que ser quisquillosos, no hemos derivado el hecho de que M sea la masa. Lo que hemos deducido (o hemos bosquejado una deducción de ello) es el hecho de que para cualquier M hay una métrica de la forma (9.5) que resolverá la ecuación de Einstein en el vacío. Una vez el hecho consumado, interpretamos M como la masa de algún objeto físico como el sol, y luego justificamos esta interpretación por comparación con el límite newtoniano o algo equivalente. Al final todo funciona, pero es importante tener en cuenta que hay conceptos que damos por sentados como evidentes y que en realidad suelen depender de nuestros compromisos teóricos previos, y que pueden sufrir cambios sutiles cuando ponemos al día nuestra teoría implícita de fondo.

DILATACIÓN DEL TIEMPO

Observemos con atención la métrica de Schwarzschild (9.5) y tratemos de comprender lo que nos está diciendo.

El interés de una métrica está en que nos permite calcular las distancias en el espacio-tiempo. Ello incluye lo referente al tiempo propio τ, que puede calcularse integrando $d\tau = \sqrt{d\tau^2} = \sqrt{-ds^2}$ a lo largo de una trayectoria de tipo temporal. Consideremos un objeto que permanece fijo en las coordenadas espaciales (r,θ,ϕ). Esto es una buena aproximación para una persona que, como usted o yo, estamos sobre la superficie de la Tierra, o incluso para la

Tierra en órbita alrededor del sol. En ambos casos se da un movimiento real (la superficie de la Tierra está rotando, y la Tierra gira alrededor del sol), pero la velocidad correspondiente es muy pequeña en comparación con la velocidad de la luz, y podemos ignorarla en la etapa inicial, al estilo del experimento mental clásico de la vaca esférica.

Para una trayectoria en unas coordenadas espaciales constantes, tenemos dr = dθ = dϕ = 0. (Esto es lo que significa estacionario: no hay incremento en esas direcciones). Si nos fijamos en (9.6), el tiempo propio a lo largo de tal trayectoria obedece a:

$$d\tau^2 = \left(1 - \frac{2GM}{r}\right) dt^2 \qquad (9.7)$$

Esto es bastante fácil de integrar. Tomemos la raíz cuadrada de ambos lados. La cantidad entre paréntesis no depende de t, así que solo es una constante, por lo que respecta al tiempo. Para un intervalo finito, tenemos:

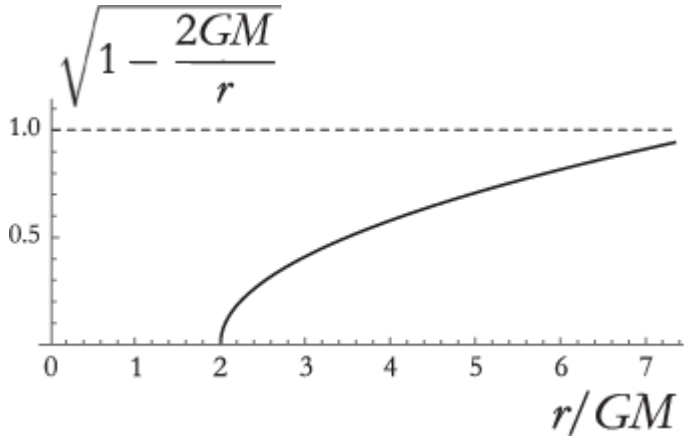
$$\Delta\tau = \int d\tau = \sqrt{1 - \frac{2GM}{r}} \int dt = \sqrt{1 - \frac{2GM}{r}} \Delta t \qquad (9.8)$$

Recuerdo que t es la coordenada para el tiempo, como mera convención humana, mientras que τ representa el tiempo propio, que es lo que miden realmente los relojes. Esta ecuación nos está diciendo que un observador estacionario experimentará una cantidad de tiempo propio proporcional al cambio en la coordenada de tiempo t, y el factor de proporcionalidad depende de la coordenada radial r.

Esta dependencia no es difícil de interpretar. Cuando r es grande, el factor $\sqrt{1-2GM/r}$ es aproximadamente 1, y el tiempo propio y el tiempo de coordenadas son el mismo. Lejos del sol, el campo gravitatorio es débil y el espacio-tiempo es casi plano. En estas circunstancias, todo reloj fijo mide

una cantidad básicamente igual al tiempo de coordenada, como Newton habría predicho.

Pero cuando r, aunque mayor, está cerca de 2GM, 2GM/r comienza a tender a 1, y √1−2GM/r tiende a cero. Así, cualquier cambio en t corresponde a un tiempo cada vez menor, a medida que r → 2GM. (Por el momento, ignore lo que sucede cuando r es menor o igual a 2GM; eso requerirá algo más de reflexión). Cuando el campo gravitatorio se hace más fuerte, el tiempo propio pasa más lentamente que el tiempo de coordenadas.



En esto consiste la dilatación gravitacional del tiempo. Hemos procurado poner de relieve que no debería decir «el tiempo va despacio en un campo gravitatorio», pero puede ver por qué es tan tentador. Los relojes siguen avanzando a un ritmo de un segundo por segundo, pero su relación con el tiempo de coordenadas ha cambiado. Esto no es relevante para un observador inmóvil, porque ¿a quién le importa una coordenada artificial?

Importa cuando comparamos lo que sucede en dos trayectorias diferentes, entre el mismo punto de partida y los puntos de llegada. Imagine que usted y un amigo suyo están lejos del sol y sincronizan sus relojes. Su amigo se queda atrás, mientras usted viaja acerándose al sol, en cuya proximidad permanece durante cierto período de tiempo, antes de regresar finalmente. Tiene cita con su amigo en algún valor del tiempo de coordenadas, pero el tiempo transcurrido de su tiempo propio será menor. Habrá envejecido menos, y su reloj señalará que ha pasado menos tiempo. (De hecho, el

campo gravitacional es extremadamente débil incluso cerca de la superficie del sol, para los criterios de la relatividad general, por lo que este efecto es muy pequeño).

Sucede lo mismo que con el gemelo del espacio de Minkowski que viajaba a una velocidad cercana a la de la luz, aunque en este caso usted se ha desplazado todo el tiempo lentamente en relación con su amigo. Todo es efecto de la curvatura del espacio-tiempo, no de su velocidad. Vea la película Interstellar, si quiere ver esto mismo representado en lo que sucede a los personajes interpretados por Matthew McConaughey y Anne Hathaway.

La dilatación gravitacional del tiempo es un efecto físico real, que ha sido verificado experimentalmente. Es un gráfico ejemplo de la idea de que la gravedad es en realidad geometría. En un sentido muy directo, la gravedad afecta el modo en que los relojes miden los intervalos en el espacio-tiempo.

SINGULARIDADES

A partir de la discusión anterior, parece que sucede algo especial cuando la coordenada radial r alcanza el valor:

$$r = 2GM \qquad (9.9)$$

Ciertamente es una magnitud importante, que se conoce como radio de Schwarzschild. Antes de emocionarnos demasiado, tengamos en cuenta que lo que sucede con el radio de Schwarzschild es completamente irrelevante para las cuestiones que nos motivaron en un principio: las relacionadas con el espacio-tiempo en la proximidad de un planeta o una estrella. El motivo es que el radio de Schwarzschild es extremadamente pequeño. Para un objeto con la masa del sol es de unos tres kilómetros, mientras que para un objeto con la masa de la Tierra es de menos de un centímetro. El sol real

tiene un radio de unos setecientos mil kilómetros, y el de la Tierra es de un poco más de seis mil kilómetros. En cualquier caso, el radio correspondiente de Schwarzschild se halla en el interior, a un nivel profundo, del cuerpo gravitatorio, no hablamos ya del espacio vacío.

La solución de Schwarzschild, por lo demás, solo se aplica a la métrica del espacio-tiempo en el vacío, en el exterior del cuerpo gravitatorio. En cuanto profundizamos en su interior, la métrica adopta una forma diferente, pero en ningún caso sucederá nada especial con r = 2GM. Más parece una curiosidad matemática, derivada de la resolución de la ecuación de Einstein en el espacio vacío, y de extrapolar sin demasiado cuidado el resultado a situaciones en que el espacio no es en absoluto un espacio vacío.

Pero nadie nos prohíbe considerar lo que suceda en una situación con un espacio vacío hasta el límite del radio de Schwarzschild, o incluso por debajo de este. ¿Qué sucedería si, debido a alguna catástrofe, la masa del sol se redujera a una esfera de apenas unos kilómetros de radio, o la masa de la Tierra se contrajera hasta un radio de menos de un centímetro? Esto nos obliga a imaginar una condensación de la materia a niveles de una densidad verdaderamente extrema, pero en astrofísica a veces suceden cosas que parecen una locura. Como veremos, tales cosas no solo pueden pasar, sino que pasan de hecho, y su resultado es un agujero negro: una zona del espacio-tiempo tan extraordinariamente curva, que la misma luz no puede escapar de ella. En el caso de un agujero negro, en que no hay materia que se interponga, el radio de Schwarzschild define el horizonte de sucesos, el punto más allá del cual es imposible regresar al mundo exterior.

Tan pronto como Schwarzschild presentó su métrica, la gente se preguntó qué iba a pasar con el radio de Schwarzschild. Si observamos la métrica (9.5), vemos que el factor $1-2GM/r$ desciende hasta cero cuando $r = 2GM$, por lo que el elemento gtt desaparece. Interpretado físicamente, esto parecería implicar que no transcurriría tiempo propio si permaneciéramos inmóviles en el radio de Schwarzschild, aunque la etiqueta de coordenadas t siguiera marcando el tiempo como de costumbre. Hmm, suena raro, pero quizá no sea demasiado preocupante; recuerda en cierto modo a lo que sucede con cualquier trayectoria nula, y la luz viaja por tales trayectorias todo el tiempo.

Más problemático es el elemento grr de la métrica, que es $1/(1-2GM/r)$ y por tanto se dispara al infinito para $r = 2GM$. Cuando una cantidad se convierte en infinita, decimos que nos encontramos ante una singularidad.

Suena mal. Pero, ¿es realmente malo? Después de todo, los elementos de la métrica dependen del sistema de coordenadas que hayamos elegido. Lo que realmente importa es la curvatura, no los elementos de la métrica. Quizás hayamos elegido un sistema de coordenadas que no era el más conveniente.

De hecho, esto es exactamente lo que sucede, aunque a los físicos les llevó bastantes años suponerlo. El aparente mal comportamiento de la métrica en el radio de Schwarzschild se debe, meramente, a una singularidad de coordenadas: no es nada que se haya definido mal desde el punto de vista de la física, sino tan solo una mala elección de coordenadas para ese punto. Cualquier función de coordenadas invariantes que pueda construir a partir del tensor de Riemann permanece finita para $r = 2GM$, aunque un elemento de la métrica se dispare. Con el radio de Schwarzschild sucede algo innegablemente interesante —constituye el horizonte de sucesos de un agujero negro, como veremos enseguida—, pero por lo que de él depende, el comportamiento del espacio-tiempo es impecable.

Esta línea de investigación nos lleva a advertir que la métrica parece también dispararse en otra ubicación, a saber, cuando $r = 0$ y por tanto $2GM/r = \infty$. En este caso, gtt se eleva a infinito y grr desciende a cero. Podríamos tener la esperanza de que suceda algo similar a lo que pasaba con el radio de Schwarzschild, y que tal vez sea nuestro sistema de coordenadas el responsable.

Esta vez no hay suerte. La ubicación $r = 0$ es una verdadera singularidad de curvatura, en que la curvatura misma del espacio-tiempo parece hacerse infinitamente grande. Esto sí, realmente es malo. Cabría esperar que esta singularidad se produjera porque hubiéramos simplificado en exceso la situación, al asumir una simetría esférica exacta, y que quizá pudiéramos evitarla en situaciones más desordenadas y realistas. Tal esperanza se vio frustrada por una serie de teoremas de singularidad, elaborados por Roger Penrose y Stephen Hawking en la década de 1960, y que demostraban que puede predecirse que las singularidades de curvatura se eleven bajo una amplia variedad de condiciones físicamente realistas.

Sin embargo, la relatividad general juega con nosotros. De acuerdo con la hipótesis de censura cósmica, formulada por Penrose en 1969, cualquier singularidad predicha por la relatividad general quedará oculta tras un horizonte de sucesos. No hay en el mundo singularidades desnudas, no cubiertas por horizontes de sucesos, que podamos estudiar de cerca. Las simulaciones numéricas modernas parecen indicar que la censura cósmica no es exactamente cierta, aunque sí casi todo el tiempo. Podemos componer unas condiciones iniciales que evolucionen hacia una singularidad desnuda, pero tales condiciones deberán ser infinitamente precisas; cualquier desviación resultaría en que la singularidad quedara encerrada tras un horizonte de sucesos. Buscar singularidades desnudas en el mundo real no parece un programa de investigación muy prometedor.

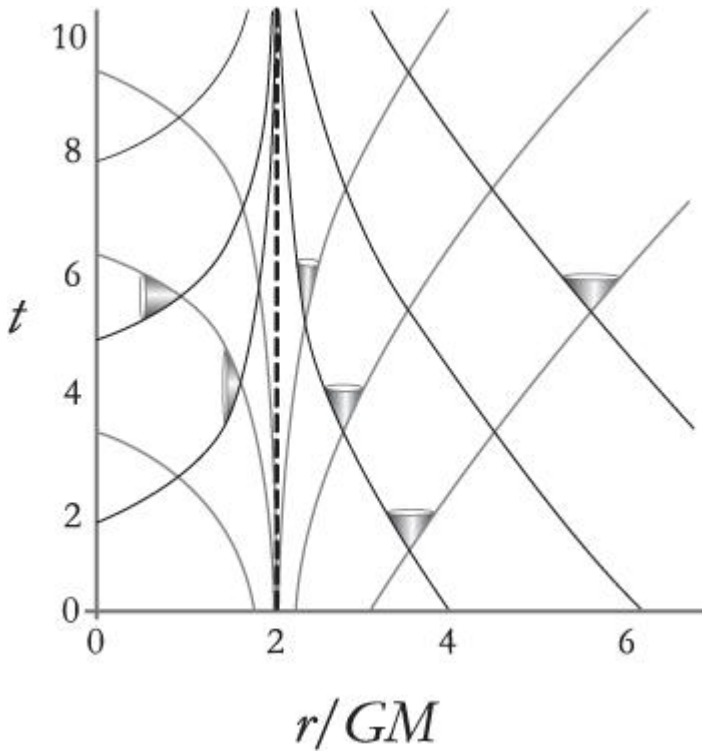
La mayor parte de los físicos piensa que las singularidades no existen realmente en la naturaleza, ni desnudas ni de ningún otro modo. Sus predicciones son una señal de que nos hemos tomado nuestra teoría demasiado en serio, en un sistema en que cabría esperar que se vulnerara. La relatividad general es una teoría clásica, después de todo, y el mundo es fundamentalmente mecánico-cuántico. Podemos albergar la esperanza de que una verdadera teoría de la gravedad cuántica alise las singularidades predichas por la teoría clásica de Einstein, o que resuelva al menos los acertijos conceptuales que suscita, pero hasta el presente no hemos llegado a comprender muy bien tan esperada teoría.

AGUJEROS NEGROS

El radio de Schwarzschild define una superficie llamada horizonte de sucesos, y la región del espacio-tiempo que está dentro del horizonte de sucesos es un agujero negro. Analicemos lo que esto significa, imaginando un espacio-tiempo descrito en todas sus partes por la métrica de Schwarzschild, sin que ningún objeto extenso, como una estrella o un planeta, se interponga.

Cuando alguien pone una métrica en nuestras manos, una buena forma de entender lo que ocurre es observar los conos de luz. Al fin y al cabo, el conjunto de todos los conos de luz es una estructura real del espacio-tiempo, no la forma preferida de nadie para definir coordenadas, ni una opción por escindir espacio y tiempo. Será esclarecedor, por tanto, representar unos cuantos conos de luz en un diagrama del espacio-tiempo de acuerdo con la geometría de Schwarzschild.

Esta es la respuesta, representada en coordenadas (t, r). Primero la contemplaremos, para justificarla después. Todo es esféricamente simétrico, lo que significa que no sucede nada especialmente interesante en las direcciones angulares θ o ϕ. Hemos dibujado rayos de luz, tanto «entrantes» como «salientes», desde unos pocos puntos, y solo hemos indicado conos de luz orientados al futuro. Si la figura es confusa, no se desanime: confundió a Einstein y a otras personas muy inteligentes durante décadas. Luego cambiaremos las coordenadas para ver mejor lo que ocurre, pero es útil pensar primero sirviéndonos de las coordenadas de Schwarzschild.



Hay muchas cosas en esta figura. A la derecha, para valores grandes de r las cosas tienen un aspecto muy normal, con rayos nulos inclinados 45 grados y los conos de luz apuntando hacia arriba. Es lógico, puesto que lejos del agujero negro no hay ningún campo gravitatorio perceptible y todo se asemeja al espacio-tiempo de Minkowski.

A medida que nos acercamos al horizonte de sucesos de r = 2GM, los conos de luz comienzan a cerrarse. Parece un poco extraño. Casi se diría que no podemos cruzar el radio de Schwarzschild porque los conos de luz no nos dejan. Claro que tal vez se deba a que nuestro sistema de coordenadas no sea el adecuado… Como pronto veremos, es exactamente eso.

Es dentro del horizonte de sucesos donde las cosas parecen volverse locas. Los conos de luz comienzan siendo muy anchos, y van estrechándose a medida que nos desplazamos hacia valores más pequeños de r. Lo más sorprendente es que apuntan a la izquierda, en lugar de hacia arriba. Pero la regla de que tenemos que permanecer dentro de nuestros conos de luz cuando avanzamos en el tiempo no ha desaparecido. Eso significa que ir en dirección a un valor más pequeño de r, dentro del horizonte de sucesos de un agujero negro de Schwarzschild, es avanzar en el tiempo.

Nos estábamos equivocando todo el tiempo con r = 0. Influidos implícitamente por nuestra intuición de un espacio-tiempo plano, lo más probable es que hayamos estado pensando en r = 0 como una ubicación en el espacio, origen de coordenadas en el centro de un agujero negro. Pero esto no es correcto, r = 0 no es una ubicación en el espacio, sino un momento en el tiempo. Es más, una vez dentro del horizonte de sucesos, es un momento que está en nuestro futuro. No podemos evitar tropezar con la singularidad, por mucho que nos esforcemos en lo contrario. Es tan inevitable como tropezarnos con el mañana.

Quizá se pregunte cómo podemos «descubrir» que r = 0 es un momento en el tiempo, más que una ubicación en el espacio. ¿Acaso no somos dueños de nuestras coordenadas? ¿No podemos definirlas como queramos? Ciertamente, pero en este caso ya hicimos nuestra elección cuando definimos r como el área de una esfera con una distancia fija con respecto a la fuente. Una vez tomada esta opción —que aceptamos con el fin de hacer que las cosas parecieran normales lejos del horizonte de sucesos—,

extendemos las coordenadas de acuerdo con un sistema diferente y nos adaptamos a lo que venga. Y lo que sucede es que r se convierte en una coordenada temporal, en lugar de espacial.

Todo esto se antoja bastante raro. Para asegurarnos de que tenemos los pies bien implantados sobre la realidad, analicemos las ecuaciones responsables de los conos de luz que hemos dibujado.

«Dibujar un cono de luz» conlleva seleccionar un punto y representar pequeños segmentos a lo largo de los cuales el intervalo espacio-temporal es cero: $ds^2 = 0$. Volviendo la mirada al elemento de línea (9.6), e ignorando θ y ϕ (puesto que no nos movemos en esas direcciones), tenemos:

$$-\left(1-\frac{2GM}{r}\right)dt^2+\frac{1}{\left(1-\frac{2GM}{r}\right)}dr^2=0 \qquad (9.10)$$

Trabajemos esto un poco. Movamos el segundo término al lado derecho, multipliquemos ambos lados por –1 para deshacernos del molesto signo general de menos, y dividamos luego entre (1–2GM/r). Obtenemos:

$$dt^2=\frac{1}{\left(1-\frac{2GM}{r}\right)^2}dr^2 \qquad (9.11)$$

Como parece que todo está elevado al cuadrado, apliquemos la raíz cuadrada a ambos lados, recordando añadir un símbolo de más/menos +/– para indicar que cualquiera de los dos signos valida una raíz cuadrada. Finalmente, dividimos ambos lados por dr, para obtener:

$$\frac{dt}{dr} = \pm \frac{1}{1 - \frac{2GM}{r}} \qquad (9.12)$$

Esto es satisfactorio, porque dt/dr es precisamente la inclinación del pequeño segmento que dibujamos para representar un cono de luz. (Recuerde que estamos considerando específicamente trayectorias nulas, no arbitrarias). El signo más/menos refleja el hecho de que la luz podría moverse hacia adentro o hacia afuera, a partir del centro.

Ahora podemos conectar esta fórmula con los conos de luz que dibujamos en la figura. Para valores grandes de r, tenemos 2GM/r ≈ 0, y por tanto dt/dr ≈ +/–1. Una inclinación de 1 o –1 indica que nuestros rayos de luz están inclinados 45 grados, que es justamente lo que sucede en el espacio-tiempo de Minkowski, y de hecho es lo que plasmamos en la figura.

A medida que r → 2GM, tenemos que 12GM/r → 0, y entonces dt/dr → +/–∞. De ahí el efecto de que los conos de luz «se cierran»: su inclinación se hace cada vez más pronunciada a ambos lados, hasta juntarse en apariencia a medida que nos acercamos al horizonte de sucesos.

Justamente al llegar al radio de Schwarzschild r = 2GM, los coeficientes de la métrica se disparan, así que dejémoslo a un lado por el momento. Pero si intrépidamente nos asomamos para echar una mirada al interior, vemos que sucede algo inesperado. Cuando r2GM, el valor de 2GM/r será superior a uno, y por tanto 12GM/r será un número negativo. Si observamos el elemento de línea (9.6), vemos que tanto gtt (el coeficiente de dt^2) como grr (el coeficiente de dr^2) cambiarán de signo. Fuera del horizonte de sucesos, gtt es negativo y grr es positivo, como en el espacio de Minkowski. Esto refleja el hecho de que t es la coordenada temporal, mientras que r es una coordenada espacial.

Pero en el interior del horizonte de sucesos, estos signos se intercambian. Lo cual tiene una consecuencia muy llamativa: ahora t es de tipo espacial y r es de tipo temporal. Físicamente, una vez estamos dentro del radio de Schwarzschild, desplazarnos hacia valores cada vez más pequeños de r no

es movernos hacia el centro, sino hacia el futuro. Que es lo que las partículas físicas hacen en la realidad, por supuesto. Y es la razón por la que es imposible evitar en último término la singularidad para r = 0.

Este comportamiento se expresa a veces (incluso por parte de personas cuyos conocimientos deberían evitarlo) como que «el tiempo y el espacio intercambian sus papeles dentro del horizonte de sucesos de un agujero negro». No. Las que intercambian sus papeles son las coordenadas t y r. Las coordenadas son una invención humana, que no deberían confundirse con características fundamentales de la realidad. El tiempo sigue siendo tiempo y el espacio sigue siendo espacio, estemos donde estemos, o sea cual sea el sistema de coordenadas que utilicemos. Si llegáramos a caer en un agujero negro lo suficientemente grande, ni siquiera advertiríamos nada especial al cruzar el horizonte de sucesos. Y desde luego nuestro reloj de pulsera no se pondría a medir distancia, en lugar de tiempo.

Unas palabras más de advertencia: si cae usted en un agujero negro, no se debata. No puede evitar la singularidad en su futuro. De hecho, tardará muy poco en toparse con él: apenas unas horas en el caso de un agujero negro de mil millones de masas solares, y en torno a una cienmilésima de segundo en el caso de un agujero negro de una masa solar. Pero eso es tan solo si cae en caída libre. Y la caída libre, recuérdelo, es un movimiento geodésico, que maximiza su tiempo propio. Si se deja llevar por un temerario impulso de escapar y trata de hacerlo, se moverá siguiendo una trayectoria de tiempo propio más corto. Desde su perspectiva, lo único que conseguirá será alcanzar antes la singularidad.

EL HORIZONTE DE SUCESOS

La singularidad en las coordenadas del radio de Schwarzschild es la única fea mancha en el cuadro que acabamos de pintar. A partir del diagrama del espacio-tiempo, como hemos mencionado, uno podría preguntarse si es posible siquiera llegar al horizonte de sucesos, puesto que los conos de luz parecen cerrarse en él. Una trayectoria de tipo temporal parecería

simplemente ascender por la coordenada t, sin llegar a cruzar el radio de Schwarzschild.

Por otra parte, al observar la dilatación gravitacional del tiempo aprendimos que cualquier cantidad fija de t se corresponde con un tiempo propio cada vez menor, a medida que nos acercamos al horizonte de sucesos. Esto es coherente con la idea de que tenemos ante nosotros un problema de coordenadas, más que físico. Tal vez, aunque parezca que una trayectoria de tipo temporal se dispara a t g+∞ , un viajero real podría atravesar en realidad una cantidad infinita de t en un tiempo propio finito. Y entonces, ¿qué sucedería con él un momento más tarde?

Lo que podría resolver este enigma es un sistema de coordenadas con mejores prestaciones para el radio de Schwarzschild. Felizmente, existen tales sistemas de coordenadas. Un sistema adecuado es el de las llamadas coordenadas de Eddington-Finkelstein, en referencia a Arthur Eddington y David Finkelstein. Este sistema se basa en las mismas coordenadas espaciales (r,θ,ϕ) que hemos utilizado para Schwarzschild, pero introduce una nueva coordenada temporal t*, definida por:

$$t^* = t + r + 2GM \log\left|\frac{r}{2GM} - 1\right| \qquad (9.13)$$

De modo que nuestra nueva coordenada de tiempo es nuestra vieja coordenada de tiempo más nuestra vieja coordenada radial más el logaritmo de cierta función de la coordenada radial. Puede parecer algo un poco arbitrario, pero existe una razón detrás de esta forma específica: cerca del horizonte de sucesos el logaritmo tiende a −∞, lo cual puede anular nuestra t que tendía a +∞, y podemos así llegar al horizonte de sucesos para un valor finito de t*.

En lugar de mostrar esto explícitamente, es más fácil formular la métrica con las coordenadas (t*,r,θ,ϕ) y observar los conos de luz. Los elementos de la métrica son:

$$g_{\mu\nu} = \begin{pmatrix} g_{t^*t^*} & g_{t^*r} & & \\ g_{rt^*} & g_{rr} & & \\ & & g_{\theta\theta} & \\ & & & g_{\phi\phi} \end{pmatrix} = \begin{pmatrix} -\left(1-\frac{2GM}{r}\right) & 1 & & \\ 1 & 0 & & \\ & & r^{22} & \\ & & & r^2(\text{sen}\theta)^2 \end{pmatrix} \qquad (9.14)$$

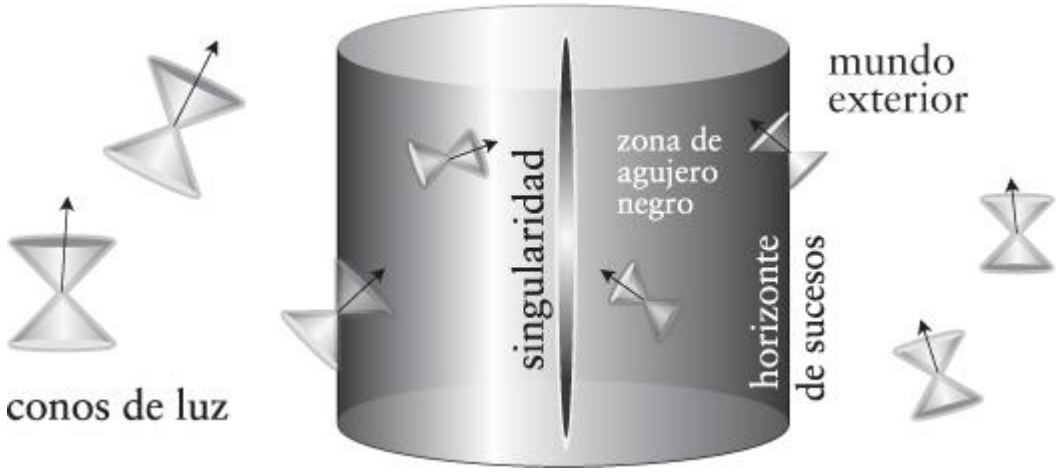
Que en forma de elemento de línea se corresponde con:

$$ds^2 = -\left(1-\frac{2GM}{r}\right)\left(dt^*\right)^2 + dt^* dr + dr dt^* + r^2 d\theta^2 + r^2(\text{sen}\theta)^2 d\phi^2 \qquad (9.15)$$

Es un poco diferente a lo que habíamos visto antes; no hay elemento grr, pero en cambio hay términos fuera de la diagonal, gt*r y grt* . Quizá le sorprende que haya desaparecido grr a pesar de no haber modificado la definición de r, pero ello se debe a que el papel de r ha sido parcialmente asumido por t*, como puede verse en (9.13).

Con las coordenadas Eddington-Finkelstein, ya no hay está presente la singularidad de coordenadas para el radio de Schwarzschild r = 2GM; todos los elementos de (9.14) son en todo momento números finitos. (El elemento llega a cero, pero el cero es un número finito). Esto se refleja en el modo en que dibujamos los conos de luz, tal como se representan en la siguiente figura (con la restitución de una dimensión angular por motivos estéticos).

En estas coordenadas, los conos de luz no se cierran en el horizonte de sucesos. Lo que hacen es inclinarse cuando nos movemos hacia decreciente r. Volvemos a ver lo mismo en el interior del horizonte; permanecer dentro de los conos de luz nos obliga a acercarnos a la singularidad para r = 0, que está en el futuro.



Estas coordenadas mejoradas nos ayudan a entender lo que sucede en el radio de Schwarzschild. El horizonte de sucesos está en una coordenada radial fija, r = 2GM, y desde el exterior el agujero negro tendrá la apariencia de una zona oscura del espacio-tiempo, con un tamaño fijo. Pero los conos de luz se alinean con el horizonte en un lado. En r = 2GM, la dirección vertical t* es en realidad nula, más que de tipo temporal (o espacial).

Esto es lo que el horizonte de sucesos tiene de especial. En r = 2GM, con el fin de permanecer en el horizonte, tenemos que movernos en una dirección específica, y hacerlo a la velocidad de la luz. Cualquier trayectoria de tipo temporal nos adentrará más profundamente en el agujero negro. Dicho de un modo más general, los caminos temporales solo pueden atravesar el horizonte de sucesos en una única dirección. No se trata de que no seamos capaces de imaginar motores para cohetes lo bastante potentes como para escapar de un agujero negro, sino que desplazarse del interior de un agujero negro al mundo exterior requeriría viajar a una velocidad superior a la de la luz.

AGUJEROS NEGROS CON CARGA ELÉCTRICA Y EN ROTACIÓN

La Tierra es esférica, pero no del todo, solo aproximadamente: es un poquitín achatada, la distancia entre los polos es más corta que el diámetro en el ecuador en un 0,3 por ciento. Además, la superficie terrestre es irregular, con sus profundos océanos y altas montañas. Como resultado, el campo gravitatorio de la Tierra no es totalmente uniforme, los satélites en órbita nos han proporcionado mapas precisos de los lugares en que la gravedad de la Tierra es un poco más intensa o un poco más débil con respecto a la media. En general, cabe decir que cada planeta tendrá sus propias características de interés particular, el campo gravitatorio de un planeta puede ser altamente distintivo.

Las cosas no son así en el caso de los agujeros negros. No son irregulares. Al contrario, se ha difundido una idea llamada teorema del no pelo, o teorema de la calvicie (aunque los matemáticos la llaman «conjetura» más que teorema, por cuanto no ha sido rigurosamente demostrada), según la cual todo agujero negro acaba adoptando un estado enteramente caracterizado por su masa, su carga eléctrica y su rotación. Dados dos agujeros negros cualesquiera con los mismos valores de estas magnitudes, tendrán los mismos campos gravitatorios. Esto significa que no va a importar qué fuera lo que interviniera en su formación: un agujero negro creado por el colapso de una estrella masiva terminará con el mismo exacto aspecto que otro agujero negro formado por una cantidad equivalente de libros de bibliotecas o de mantequilla de cacahuete. Estos últimos casos probablemente no sucedan en la naturaleza, pero si ocurriesen, no podríamos deducirlo examinando el agujero negro resultante.

Los agujeros negros con carga eléctrica pero sin rotación estarán rodeados por un campo eléctrico esférico y simétrico, como lo estaría una partícula con carga eléctrica. Ese campo eléctrico contiene energía, que deberá tenerse en cuenta cuando se resuelva la ecuación de Einstein. La métrica de tal agujero negro la proporciona la solución de Reissner-Nordström, obtenida por diversas personas poco después de la solución de Schwarzschild. No es demasiado sorprendente: puesto que todo sigue siendo esféricamente simétrico, la solución no puede ser muy diferente de la de Schwarzschild.

Al menos exteriormente. La solución de Reissner-Nordström, tomada al pie de la letra y llevada a sus últimas consecuencias, describe una serie infinita de agujeros negros que conectan universos separados, todos ellos ocultos tras el horizonte de sucesos.

Por eso no siempre hay que buscar soluciones exactas a las ecuaciones y tomarlas al pie de la letra, especialmente cuando uno trata de llevar su aplicación más allá de la situación física para cuya descripción se habían concebido. Incluso la solución de Schwarzschild, llevada al extremo, describe dos universos diferentes conectados por un agujero de gusano. Este hecho fue advertido por Ludwig Flamm en 1916 y redescubierto por Einstein y su colaborador Nathan Rosen en 1935, por lo que el agujero de gusano de Schwarzschild se conoce a veces como puente de Einstein-Rosen.[2] Pero nadie cree que haya agujeros negros reales, diseminados por ahí en nuestra galaxia, que oculten agujeros de gusano y universos extra tras sus horizontes de sucesos. La causa es que los agujeros negros reales no pueden concebirse precisamente como espacios vacíos (o campos eléctricos) en toda su extensión. Están formados por materia que cae en su interior, y la existencia de esa materia afecta a la métrica del espacio-tiempo de forma importante. Como resultado, los agujeros negros reales poseen singularidades en su interior, pero no puertas a otros universos.

Costó mucho más tiempo encontrar una métrica para un agujero negro en rotación. Fue finalmente elaborada por Roy Kerr en 1963, por lo que se conoce en la actualidad como solución de Kerr. La dificultad estriba en la falta de simetría esférica; lo que existe es una dirección privilegiada, el eje de rotación del agujero negro. Pero la labor de Kerr supuso algo más que una proeza matemática. En el mundo real, no esperamos que los agujeros negros tengan una carga eléctrica sustancial; si alguno la tuviera, sería rápidamente neutralizada por la atracción y absorción de partículas de carga opuesta. Pero sí cabe esperar que los agujeros negros roten, y que lo hagan a gran velocidad. Casi todos los agujeros negros del universo podrían ser descritos con precisión por la métrica de Kerr.

EL TEOREMA DEL ÁREA

Aunque la idea de agujero negro estaba implícita en el trabajo de Schwarzschild desde 1915, los físicos tardaron mucho tiempo en comprender el concepto. Esta dilación se debió en gran parte a la dificultad en apreciar qué características eran físicamente reales y cuáles eran meramente accidentales, producto de las coordenadas, motivo por el cual hemos insistido tanto en esta diferenciación. Se entendía, por ejemplo, que la dilatación temporal gravitatoria aumentaba infinitamente para un observador que permaneciera justamente en el radio de Schwarzschild, por lo que el tiempo parecería congelarse desde la perspectiva de un observador exterior. Por tanto, razonaban, un objeto que se redujera a tal tamaño se convertiría en una «estrella congelada», atorada por siempre en ese radio. No se preguntaban con detenimiento, sin embargo, qué sucedería desde el punto de vista de un observador que cayera al otro lado, el cual, como ahora sabemos, se zambulliría de lleno en un agujero negro.

Fue a finales de la década de 1950 cuando finalmente se llegó a comprender los agujeros negros, gracias al trabajo de David Finkelstein, Martin Kruskal y otros. Esto contribuyó a la inauguración de una nueva era de activa investigación en torno a la relatividad general, encabezada por John Wheeler en los Estados Unidos, Dennis Sciama en el Reino Unido y Yakov Zeldovich en la Unión Soviética. Wheeler popularizó el término «agujero negro», cuya investigación fue propulsada por Penrose, Hawking y sus colaboradores en las décadas de 1960 y 1970.

Un importante fruto de aquel período fue el teorema del área, demostrado por Hawking en 1971. Determina que el área del horizonte de sucesos de un agujero negro, o el área combinada de varios agujeros negros, con el tiempo solo se incrementa, nunca decrece. Como todo buen teorema, se fundamenta en supuestos. Suponemos, por ejemplo, que las partículas reales solo tienen masa positiva, de lo contrario podríamos arrojar partículas de masa negativa a un agujero negro para hacer que este se redujera. Y por supuesto estamos pensando puramente con mentalidad clásica: una vez entre en juego la mecánica cuántica, los agujeros negros perderán energía en forma de radiación, por lo que podrán reducirse, a fin de cuentas, aunque por lo general suceda de forma extremadamente lenta.

El teorema del área cobra importancia cuando pensamos en dos agujeros negros que se funden. Cuando esto sucede, parte de la energía total se perderá en forma de ondas de radiación gravitatoria, ya que el espacio-tiempo resulta sacudido por los agujeros negros en rápido movimiento. Pero el teorema del área nos asegura que no escapa demasiada energía. El área del horizonte de sucesos del agujero negro resultante de la fusión será como mínimo igual de grande que la suma de las áreas de los dos agujeros negros originarios.

Quizá se pregunte por qué no decimos simplemente que la masa de un agujero negro nunca decrece, en lugar de decir que nunca decrece el área de su horizonte de sucesos, dado que el radio de Schwarzschild es $r = 2GM$ y el área del horizonte es por tanto $A = 4\pi r^2 = 16\pi G^2M^2$. La respuesta es que la masa puede decrecer, y ello se debe a que existen los agujeros negros en rotación.

Para un agujero negro en rotación, el área del horizonte de sucesos depende tanto de la masa como de la cantidad de rotación. Resulta que de un agujero negro en rotación puede extraerse energía arrojando en su interior materia con un momento angular opuesto. Roger Penrose concibió explícitamente un modo de hacerlo, que se conoce como proceso de Penrose. Podríamos imaginar una civilización extraterrestre avanzada que alimentara su red eléctrica extrayendo energía de un agujero negro supermasivo en rotación. Sin embargo, el teorema de Hawking nos garantiza que el área del horizonte de sucesos seguirá incrementándose mientras tanto, y que al final el agujero negro dejará de girar y no quedará más energía que extraer.

MECÁNICA DE LOS AGUJEROS NEGROS

Hay algo en el teorema del área que debería preocuparle. Dice que, con el paso del tiempo, el área del horizonte de sucesos de un agujero negro no hará sino crecer. Pero esto implica la presencia de una flecha del tiempo: el área se incrementa en dirección al futuro, no en dirección al pasado. ¿De

dónde surge ahora esto? La ecuación de Einstein no trata el pasado y el futuro de una manera en modo alguno diferente.

Parte de la respuesta es limitada y técnica al mismo tiempo. Mientras que la relatividad general no encierra en sí misma una flecha del tiempo, las soluciones en torno a los agujeros negros que hemos considerado sí lo hacen. En un segundo plano de nuestra mente, imaginamos que hubo en el pasado una estrella o algún otro objeto astrofísico que colapsó (mientras el tiempo avanzaba hacia el futuro) para crear nuestro agujero negro. Sería posible reproducir toda nuestra discusión matemática con la dirección del tiempo inversa. El resultado sería un agujero blanco, que contendría una singularidad en el pasado, rodeado por un horizonte de sucesos del que la materia podría escapar, pero no regresar. Un agujero blanco es un agujero negro retrocediendo en el tiempo. No pensamos que los agujeros blancos existan en la naturaleza, pero la cosmología moderna postula que nuestro universo observable surgió de una singularidad de tipo Big Bang acaecida en el pasado, y en todo ello existe ciertamente una semejanza de familia. El universo, considerado en su totalidad, viene a ser como un agujero blanco.

Esta interconexión sugiere una relación más profunda entre los agujeros negros y la termodinámica (que es, después de todo, el origen de la flecha del tiempo). A comienzos de la década de 1970, los físicos señalaban que se daba una correspondencia intrigante entre el teorema de Hawking («el área siempre aumenta») y la segunda ley de la termodinámica («la entropía siempre aumenta»). Se consideró una analogía graciosa, nada más.

Es decir, hasta la llegada de los trabajos de Jacob Bekenstein, un estudiante graduado que colaboraba con John Wheeler. Bekenstein sugirió que en los agujeros negros sí se da la entropía en realidad, y que tal entropía es proporcional al área del horizonte de sucesos. Los físicos más famosos se rieron de él, y con buen motivo: si en los agujeros negros hubiera entropía, entonces, por las reglas convencionales de la termodinámica, en ellos debería haber también temperatura. (En la termodinámica clásica, cuando cambiamos la energía de un sistema de volumen fijo, el cambio de energía dividido por el cambio de entropía es igual a la temperatura). Y eso significa que desprenderían radiación. Y eso significa que no serían negros, a fin de cuentas.

Hawking, en particular, se molestó con la propuesta de Bekenstein. Pero la tomó lo suficientemente en serio como para aplicarse en demostrar su falsedad, utilizando técnicas que combinaban la teoría cuántica de campos y la relatividad general. Como hoy día es bien sabido, lo que terminó demostrando es que Bekenstein tenía razón. En los agujeros negros hay entropía, y también temperatura, y desprenden radiación, una vez considerados los efectos cuánticos. Pero esta historia la dejaremos para otra ocasión, una vez tengamos asimilado algo de mecánica cuántica. Por ahora nos contentaremos con saber que la llamada radiación de Hawking es extremadamente tenue, en verdad. Para un agujero negro de una masa solar, la temperatura es inferior a una millonésima de grado Kelvin, muy por debajo del nivel al que pudiera ser detectable (y los agujeros negros más grandes tienen temperaturas aún más bajas).

EL MUNDO REAL

No dejó de causar como mínimo cierta sorpresa entre los físicos cuando el Premio Nobel de Física de 2020 se concedió a Roger Penrose, Reinhard Genzel y Andrea Ghez. La inclusión de Penrose, en particular, representó un cambio de orientación para la Real Academia de Ciencias de Suecia. Nadie puede dudar de la brillantez ni de la importancia de su trabajo, pero los premios Nobel suelen concederse con mayor frecuencia por descubrimientos experimentales o por modelos teóricos muy específicos, y no por demostrar que una teoría establecida (la relatividad general) implica la existencia de un fenómeno particular (los agujeros negros). Pero el caso de Penrose se vio enormemente reforzado por el trabajo de Genzel y Ghez, que habían reunido sólidas pruebas observacionales acerca de la existencia de un agujero negro real en el centro de nuestra galaxia.

Los agujeros negros han pasado de ser curiosidades teóricas a ocupar un puesto en la vanguardia de la astrofísica moderna. No hemos visto ninguno de cerca —quizá sea mejor así—, pero hay pruebas abrumadoras de que existen y desempeñan un papel de gran importancia en múltiples procesos astrofísicos.

De hecho, hay toda una colección de agujeros negros, de diferentes tamaños y con historias que nos cuentan orígenes diversos. Quizás el tipo de agujero negro más famoso es el que categoriza a los formados al final de la vida de una estrella masiva. La energía estelar procede de la fusión de elementos ligeros en otros más pesados en el núcleo de la estrella. Al final del proceso, la estrella agotará todo su combustible utilizable. En esa etapa, la mayoría de las estrellas se estabilizarán en forma de enana blanca, que irá enfriándose gradualmente con el tiempo. Otras estrellas más masivas se contraerán hasta formar una estrella de neutrones, en la cual protones y electrones se habrán combinado constituyendo en neutrones. Las estrellas más masivas, sin embargo, colapsarán hasta convertirse en agujeros negros. Se piensa que la mayor parte de los agujeros negros formados de este modo tendrán como mínimo tres veces la masa de nuestro sol.

Cabe preguntarse cómo vamos a poder detectar la existencia de agujeros negros atribuidos a remanentes estelares. Al fin y al cabo son negros. Pero también giran, y hasta una pequeña cantidad de rotación en la estrella inicial puede ser significativa cuando la masa queda reducida al tamaño de un agujero negro. Como resultado, la materia atraída al agujero negro tiende a acumularse en forma de disco de acrecimiento en su plano ecuatorial. Puede haber una considerable cantidad de esta materia, especialmente si el agujero negro forma parte de un sistema binario, junto con otra estrella. La temperatura de la materia en un disco de acrecimiento llega a ser muy alta, lo suficiente como para emitir radiación de rayos X de alta intensidad. Son estos rayos X los que pueden observar los astrónomos. La mayoría de agujeros negros no van a estar rodeados por discos de acrecimiento adecuados para la observación, pero, dado el número de estrellas masivas de la Vía Láctea, los astrónomos estiman que podría haber cientos de millones de agujeros negros de masa estelar dispersos por nuestra galaxia. (En la cual hay unos cien mil millones de estrellas, por lo que los agujeros negros siguen representando una muy pequeña fracción).

Hay una población totalmente diferente, los agujeros negros supermasivos, acechando en el centro de las galaxias. Su masa es millones o miles de millones de veces mayor que la del sol, y los astrónomos creen que están presentes en la mayoría de galaxias mayores del universo. Nuestra propia Vía Láctea cuenta con un agujero negro de unos cuatro millones de masas

solares. Hemos tenido conocimiento de ello gracias en parte al seguimiento de las órbitas de alta velocidad de las estrellas que giran alrededor de una región compacta y oscura localizada en la constelación de Sagitario. Por estas observaciones recibieron Genzel y Ghez su Premio Nobel compartido.

La Vía Láctea es una galaxia madura, cuya mayor parte de gas y polvo se ha convertido en estrellas. No ha quedado por tanto mucha materia perdida para caer en el agujero negro central, por lo que el nuestro es relativamente tranquilo. Pero las galaxias más jóvenes a menudo cuentan con agujeros negros dotados de discos de acrecimiento enormes y de un resplandor muy brillante. Los vemos dispersos por el universo observable en forma de cuásares o (más generalmente) de núcleos galácticos activos.

Kip Thorne nos cuenta la historia del Primer Simposio de Texas sobre Astrofísica Relativista, celebrado en Dallas en diciembre de 1963.[3] El astrónomo Maarten Schmidt acababa de medir por primera vez la distancia a un cuásar, demostrando que está extremadamente lejos. Dado lo brillante que lo vemos desde la Tierra, tal objeto tiene que ser extraordinariamente luminoso. Los astrónomos presentes en el simposio bullían de emoción, disparando una tras otra hipótesis acerca de la naturaleza de los cuásares y la importancia que podría tener la relatividad para entenderlos. Entretanto, un joven matemático de Nueva Zelanda impartió una esotérica charla de diez minutos sobre una nueva solución a la ecuación de Einstein para un espacio-tiempo en rotación. La audiencia la ignoró en gran medida, muchas personas aprovecharon para salir de la sala y hacer un descanso. El ponente era Roy Kerr, y casi nadie se dio cuenta de que acababa de describir la métrica para un agujero negro en rotación, que iba a revelarse como el ingrediente crucial para entender los cuásares.

Desde 2015, contamos con una nueva forma de obtener información sobre los agujeros negros: las ondas gravitatorias. Si la gravedad es la curvatura del espacio-tiempo, una onda gravitatoria es una ondulación en la curvatura que se propaga a la velocidad de la luz. Así como las ondas electromagnéticas ordinarias pueden generarse por movimientos rápidos de partículas con carga eléctrica, las ondas gravitacionales por movimientos rápidos de objetos masivos.

El problema es que la gravedad es una fuerza débil, por lo que tales ondas son difíciles de detectar. En 2015, el Observatorio de Ondas Gravitatorias por Interferometría Láser (LIGO, por sus siglas en inglés), que trabaja en colaboración con el Observatorio Virgo en Europa, anunció haberlas detectado por primera vez. LIGO consta de dos observatorios, cada uno de ellos con un par de tuberías de cuatro kilómetros de longitud orientadas en ángulo recto. Los rayos láser viajan a través de las tuberías y rebotan en los espejos de los extremos. Una onda gravitatoria induce a su paso una pequeña distorsión del espacio-tiempo a lo largo de las tuberías, alterando el tiempo que tardan los rayos láser en viajar a los espejos y volver. La señal es realmente pequeñísima: una onda típica causará una oscilación en el reflejo de los espejos de una distancia inferior a la anchura de un protón. No es de extrañar que se hayan necesitado tantos años de trabajo de diseño y construcción, y tantos millones de dólares, para conseguir implantar un instrumento tan sensible. Tampoco fue ninguna sorpresa que en 2017 se concediera el Premio Nobel a Rainer Weiss, Kip Thorne y Barry Barish por su trabajo para el desarrollo de LIGO.

El suceso detectado en 2015 fue el resultado de la fusión de un agujero negro de 36 masas solares con otro de 29 masas solares. Los dos colosos, situados a unos mil millones de años luz de distancia, llevaban orbitando estrechamente el uno alrededor del otro un periodo de tiempo desconocido. El movimiento orbital había producido ondas gravitatorias, y los agujeros negros habían ido acercándose cada vez más, a medida que el sistema perdía la energía empleada en esa radiación. La coalescencia final fue cuestión de segundos, el hecho en sí mismo duró tan solo hasta que el agujero negro resultante se estabilizó rápidamente (de acuerdo con el teorema del no pelo).

Desde entonces, LIGO y Virgo han detectado docenas de episodios de este tipo. En la mayoría de ellos intervenían agujeros negros de entre diez y cien veces la masa del sol. Esto depende en parte de los tipos de agujeros negros existentes en el universo, pero en parte también de las longitudes de onda específicas a las que son sensibles los detectores.

Los astrofísicos modernos están entusiasmados ante las perspectivas de lo que podríamos aprender de esta nueva ventana que se ha abierto al cosmos.

Sabremos más sobre los agujeros negros, por supuesto, pero también, potencialmente, sobre los ciclos de vida de las estrellas, la estructura de las galaxias y quizá sobre el tamaño y la forma del universo. Y por encima de todo, como todo buen científico, estaremos preparados para lo que pudiera ser, por inesperado y sorprendente que sea.

1 Esta elección significa que r no será necesariamente (o realmente, como resulta) la distancia al origen. Es una cantidad física diferente, y no sabemos de antemano cómo se relacionará con la distancia medida a lo largo de esferas de un radio dado.

2 Me pidieron que participara como asesor científico para la película Thor , de 2011. Si le llama la atención Jane Foster, interpretada por Natalie Portman, hablando de puentes de Einstein-Rosen, es culpa mía.

3 Kip Thorne, Black Holes and Time Warps: Einstein's Outrageous Legacy. W. W. Norton, Nueva York, 1994, cap. 9. (Agujeros negros y tiempo curvo . Crítica. Barcelona, 2010).

APÉNDICES

APÉNDICE A:

FUNCIONES, DERIVADAS E INTEGRALES

En este apéndice repasamos algunas funciones muy comunes y algunas de sus aplicaciones, por si le apetece probar de resolver algunas de las ecuaciones de que hemos hablado.

Un rápido comentario sobre la notación. Solemos utilizar letras del final del alfabeto, como x, y, z, para indicar variables (cantidades que permanecen indeterminadas de una ecuación, y que se resolverán o no más tarde). Las letras del comienzo del alfabeto, como a, b, c, es costumbre reservarlas para las constantes (cantidades que tienen un valor determinado, aunque puede ser que no digamos, o no sepamos, cuál es). Y las letras del final de primer cuarto del alfabeto, como f y g, suelen usarse para las funciones (correspondencias entre variables). Así, una notación estándar sería: $f(x) = ax + b$, donde x es una variable, a y b son constantes y $f(x)$ es una función de x. Si x es una variable, significa que la relación aludida con la letra f se supone que valdrá para cualquiera que sea el valor de x, mientras que concebimos a y b como cantidades fijas, aunque no digamos qué cantidades son. Se trata de una distinción sutil.

Por supuesto, todo esto es una cuestión meramente convencional, y en principio podemos utilizar las letras que queramos. Pronto nos quedaremos sin alfabeto y tendremos que recurrir a letras griegas.

INTEGRALES DEFINIDAS E INDEFINIDAS

En nuestra introducción a las integrales, en el Capítulo 2, pasábamos por alto un detalle importante: una integral representa el área bajo una curva. Pero esto solo tiene sentido si especificamos dónde comienza y dónde termina la región cuya área estamos describiendo. Esto nos lleva a distinguir entre integral definida, aquella en que definimos los puntos límite, e integral indefinida, aquella en que dejamos tales puntos sin especificar.

Consideremos un caso en que la integral de una función f (x) es otra función F (x). En otras palabras:

$$\int f(x)dx = F(x) \qquad \text{(A.1)}$$

Esto es una integral indefinida. Está claro que hemos sido un poco descuidados, puesto que no hemos especificado los puntos extremos. A veces, para ser más cuidadosos, añadimos: «más una constante», para indicar que el valor exacto dependerá de la región elegida.[1] Pero normalmente damos por sentado que usted se mantiene alerta y que sabe lo que hay que hacer con la expresión para la integral. A lo largo de estos libros, por «integral de una función» nos referimos en general a una integral indefinida.

La integral definida nos permite delimitar la región de integración de forma explícita, especificando el inicio en la parte inferior del signo de integral y el punto final en la parte superior:

$$\int_a^b f(x)dx = F(b) - F(a) \qquad \text{(A.2)}$$

De modo que la integral definida es la diferencia entre el valor de la integral definida en el punto final y su valor en el punto inicial. Veámoslo en la práctica.

FUNCIONES CONSTANTES

Consideremos una función extremadamente simple, es decir, una constante, f (x) = c. No es que estemos entrando muy en detalle, pero se trata de empezar poco a poco. La inclinación de una constante es plana, así que no es ninguna sorpresa que su derivada sea cero:

$$\frac{d}{dx}c = 0 \qquad \text{(A.3)}$$

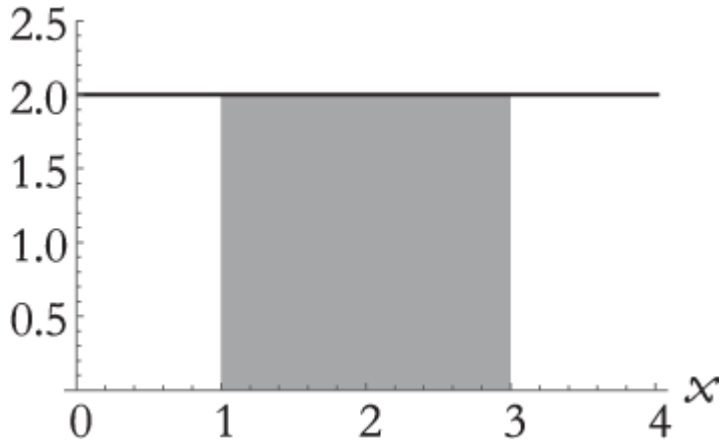
La integral indefinida es proporcional a la propia x:

$$\int c\,dx = cx \qquad \text{(A.4)}$$

Esto significa que la integral definida es proporcional a la distancia entre los dos puntos extremos:

$$\int_a^b c\,dx = c(b-a) \qquad \text{(A.5)}$$

Esto se ejemplifica en la figura para c = 2, a = 1, b = 3. El área bajo la curva es 2x(3−1) = 4, como cabría esperar.



Los paréntesis de (A.5) significan que la cantidad c se multiplica por la cantidad (b−a), no que b−a sea el argumento de una función, como lo es x en f (x). Notación similar, significados diferentes; se supone que usted es capaz de discernir la diferencia a partir del contexto.

COMBINACIONES LINEALES

En matemáticas, decimos que una suma con la forma af (x) + bg (x), donde a y b son constantes, es una combinación lineal de las dos funciones f (x) y g (x). La palabra «lineal» hace referencia al hecho de que cada función aparece tan solo una vez, y elevada a la primera potencia; al multiplicar funciones entre sí, o elevarlas a una potencia diferente, ya no serían lineales.

Tanto la diferenciación como la integración son operadores lineales. Esto significa que la derivada de una combinación lineal es la combinación lineal de las derivadas correspondientes, y del mismo modo para las integrales. Así pues, para las derivadas tenemos:

$$\frac{d}{dx}\left[af(x)+bg(x)\right]=a\frac{df}{dx}+b\frac{dg}{dx} \qquad \text{(A.6)}$$

Y para las integrales:

$$\int \left[af(x) + bg(x) \right] dx = a \int f(x) dx + b \int g(x) dx \qquad \text{(A.7)}$$

Esto funciona (por supuesto) incluso si el segundo término está completamente ausente y la función que queremos diferenciar o integrar es simplemente af (x). En estos casos podemos sin más «sacar la constante fuera de la derivada (o de la integral)». Y si estamos integrando con respecto a x (como se indica con el símbolo dx), entonces cualquier cosa que no dependa de x cuenta como constante, aun si depende de otra variable:

$$\int f(x) g(y) dx = g(y) \int f(x) dx$$

PRODUCTOS

Supongamos que tenemos el producto de dos funciones, f (x) g (x), donde por tanto podemos omitir (x) y escribir simplemente fg para mayor limpieza. A partir de aquí, existe una fórmula sencilla aunque poco intuitiva para la derivada del producto:

$$\frac{d}{dx} fg = f \frac{dg}{dx} + g \frac{df}{dx} \qquad \text{(A.8)}$$

Podría expresarse esto como «la primera por la derivada de la segunda, más la segunda por la derivada de la primera». Se conoce como la regla de

Leibniz, por Gottfried Leibniz.

La existencia de tan elegante procedimiento para calcular la derivada de un producto es la razón implícita por la que los matemáticos consideran generalmente «fácil» la diferenciación. Casi todas las funciones que nos interesen pueden elaborarse (quizá recursivamente) en forma de combinaciones de otras funciones por adición, multiplicación, y así sucesivamente. La regla de Leibniz implica que las derivadas de la mayor parte de las funciones pueden expresarse explícitamente en forma de otras funciones (o «en forma cerrada», como decimos nosotros).

Sería lo más natural presentar en este punto la fórmula para la integral de un producto de dos funciones, pero por desgracia no existe tal fórmula. La integración es difícil, tanto conceptualmente como en la práctica.

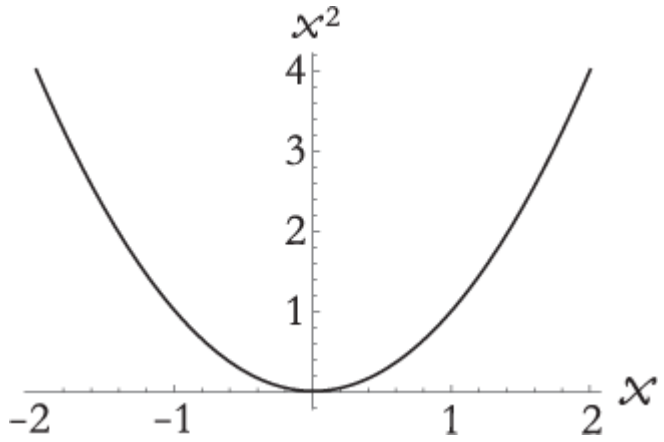
POTENCIAS

Si vamos de principios generales a funciones más específicas, lo más común que encontraremos es una variable x elevada a una potencia a, expresado como xa. Aquí la variable es la base y la potencia es el exponente, pero hay que distinguir esto de la función exponencial, en que una constante está elevada a la potencia de la variable, como hablaremos más abajo. Si a es un entero positivo, xa es igual a multiplicar x por sí mismo a veces. Pero no hay problema en hacer un poco de trampa matemática y definir xa no siendo a un número entero, siendo negativo o incluso siendo un número complejo.

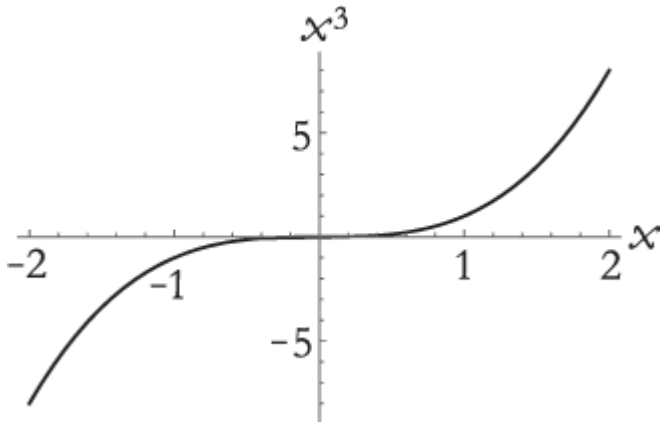
Dos cosas útiles sobre potencias: cuando multiplicamos potencias de una misma variable, simplemente sumamos los exponentes, y cuando elevamos una potencia a otra potencia, los exponentes se multiplican.

$$\left(x^a\right)^{\frac{1}{a}} = x^{\left(\frac{a}{a}\right)} = x^1 = x \qquad \text{(A.9)}$$

Veamos algunos ejemplos sencillos (y probablemente familiares). La función x^2 es una parábola.



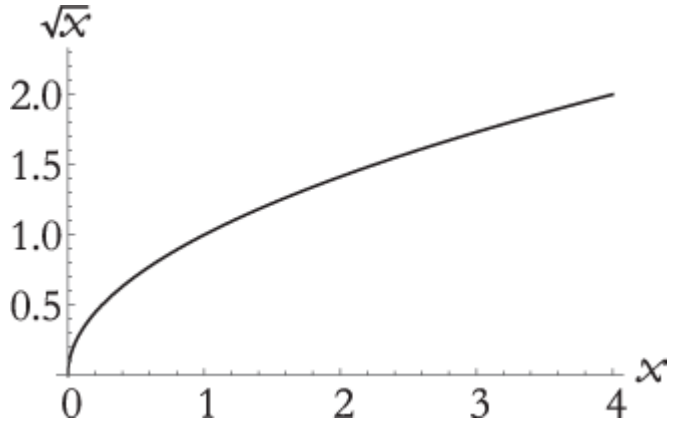
El valor de esta función nunca es negativo, porque al multiplicar dos números negativos entre sí (x por sí mismo) el resultado es un número positivo. Lo mismo sucede siempre que elevemos x a cualquier otro entero par, y la curva tendrá un aspecto cualitativamente similar. Cuando la potencia a es un entero negativo, el lado negativo de la función es menos el lado positivo, como en el caso de x^3.



Podemos también definir potencias en forma de fracciones, aunque en este caso estamos restringidos a hacerlo con valores no negativos para x. Conceptualmente, cuando elevamos algo a la potencia 1/a, deshacemos la acción de elevarlo a la potencia a, porque las sucesivas potencias se multiplican en el exponente:

$$\left(x^{a}\right)^{\frac{1}{a}} = x^{\left(\frac{a}{a}\right)} = x^{1} = x \qquad \text{(A.9)}$$

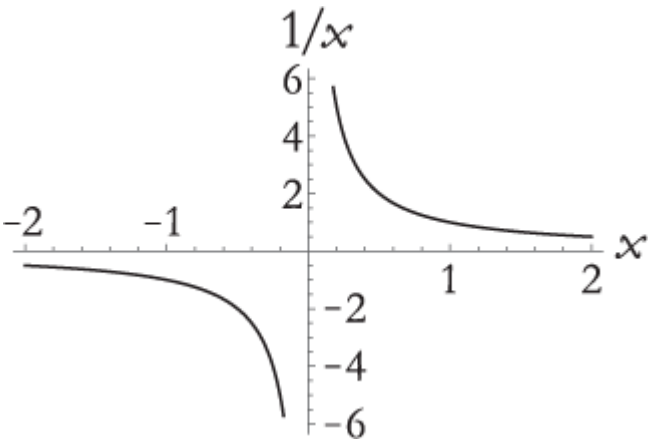
Como resultado, x½ = √x tiene el aspecto de una parábola inclinada hacia un lado.



Potencias negativas o inversas pueden igualmente definirse, considerándolas como el producto de una potencia y su negativo. La función recíproca x−1 = 1/x, por ejemplo, puede definirse exigiendo que satisfaga:

$$x \cdot x^{-1} = x^{1-1} = x^{0} = 1 \qquad \text{(A.10)}$$

La gráfica resultante tiene una discontinuidad en x = 0, pero no hay nada que temer. Consideramos simplemente que 1/x no está definido en ese punto.



La derivada de una potencia es la simplicidad misma: se baja la potencia en uno, y se multiplica la expresión entera por la potencia originaria:

$$\frac{d}{dx}x^a = ax^{a-} \qquad \text{(A.11)}$$

La integral, bastante sensatamente, aumenta la potencia en uno:

$$\int x^a dx = \frac{x^{a+1}}{+1} \qquad \text{(A.12)}$$

Como entretenimiento, puede comprobar que si primero calcula la derivada y luego la integral, vuelve a obtener la función originaria, como debe ser.

Existe un problema latente, no obstante: cuando a = −1, parece que estaríamos dividiendo por cero en (A.12). Ciertamente así es, y por eso hay

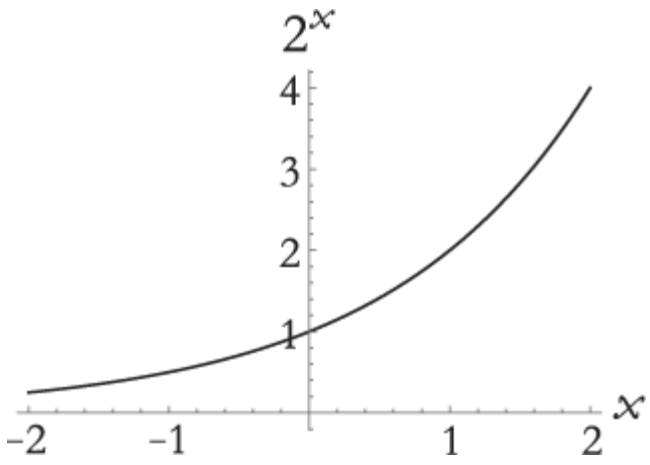
una fórmula especial para un caso así:

$$\int x^{-1} dx = \ln|x| \qquad \text{(A.13)}$$

Las barras verticales de |x| son signos de valor absoluto: si x es positiva, la dejamos tal cual; si x es negativa, la multiplicamos por −1, y así el resultado es no negativo. La función lnx es el logaritmo natural, así que es lo siguiente que vamos a explicar.

FUNCIONES EXPONENCIALES Y LOGARITMOS

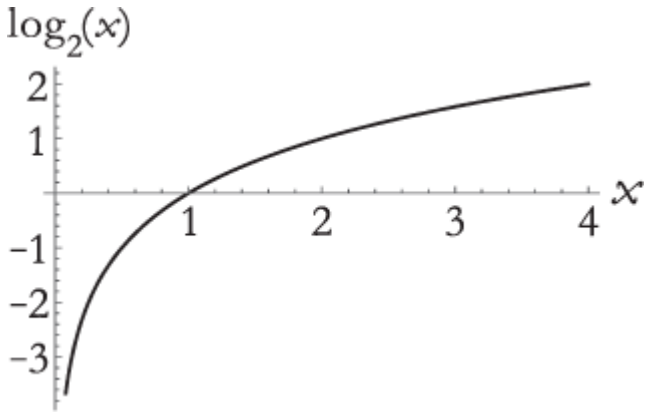
Pasaremos ahora al caso en que la variable está en el exponente, que se conoce como función exponencial, f (x) = ax. La idea básica es bastante clara, como se muestra en la figura para a = 2:



La función inversa de ax es el logaritmo (con base a), que cumple:

$$\log_a\left(a^x\right)=a^{\log_a(x)}=x \qquad \text{(A.14)}$$

Al igual que las funciones exponenciales son funciones paradigmáticas de aumento rápido, los logaritmos son funciones paradigmáticas de aumento lento, a medida que x es mayor. El logaritmo de 1 es igual a 0, y loga(a) = 1. Para una x muy pequeña, el logaritmo tiende a −∞, lo cual tiene sentido si pensamos en loga(x) como «la base a la que tendríamos que elevar a para obtener x». Dado a, para obtener un número cercano a cero elevándolo a alguna potencia, tal potencia tiene que ser grande y negativa.



Existe una noción muy particular en el mundo de las funciones exponenciales y los logaritmos. Se trata de una cantidad que desempeña un papel especial: el número de Euler.

$$e=2.71828\ldots \qquad \text{(A.15)}$$

La expresión de e contiene un número infinito de términos que no llegan a adoptar nunca un patrón repetitivo. Al igual que π = 3,14139…, es irracional, un número no expresable como razón entre dos enteros. Hay muchas formas de definir e, pero quizá la más satisfactoria sea decir que ex es la única función no constante que es igual a su propia derivada:

$$\frac{d}{dx}e^x = e^x \qquad \text{(A.16)}$$

Para funciones exponenciales con otras bases, la derivada toma la forma:

$$\frac{d}{dx}a^x = \ln(a)a^x \qquad \text{(A.17)}$$

Hemos definido aquí el logaritmo natural, que es simplemente el logaritmo con base e:

$$\ln(x) = \log_e(x) \qquad \text{(A.18)}$$

Es el logaritmo natural que aparece en (A.13), la integral de 1/x. Observando (A.17), y recordando que loga(a) = 1 para cualquier a, vemos que el molesto factor ln(a) desaparece cuando a = e, dejándonos la elegante forma (A.16). Por esta razón la mayor parte de los físicos utilizan e como base de sus logaritmos siempre que es posible.

La fórmula de la integral para una función exponencial es igualmente clara:

$$\int a^x dx = \frac{a^x}{\ln(a)} \qquad \text{(A.19)}$$

La derivada de un logaritmo es:

$$\frac{d}{dx}\log_a(x)=\frac{1}{\ln(a)}x^{-1} \qquad (A.20)$$

Y su integral es:

$$\int \log_a(x)dx=\frac{x\ln(x)-x}{\ln(a)} \qquad (A.21)$$

Puede apreciar cómo adquieren su más bonito aspecto cuando a = e y ln(a) = lne = 1.

FUNCIONES TRIGONOMÉTRICAS

El último grupo de funciones famosas que vamos a considerar es el de las funciones trigonométricas, en especial el seno y el coseno. En este caso, el argumento de la función será generalmente un ángulo, en lugar de un número real, por lo que lo denotaremos mediante en lugar de x. Y lo que es importante, mediremos nuestros ángulos en radianes, mejor que en grados. Ciento ochenta grados se corresponde con radianes, de modo que es sencilla la conversión de una medida a otra.

En el Capítulo 3 ya introdujimos las funciones trigonométricas, por lo que saltaremos a algunas de sus interesantes propiedades. El teorema de Pitágoras nos proporciona de entrada una famosa relación entre el seno y el coseno:

$$(\text{sen}\theta)^2+(\cos\theta)^2=1 \qquad (A.22)$$

Si tenemos un vector con componentes i en un espacio euclidiano plano y tridimensional, podemos recurrir una vez más al teorema de Pitágoras para definir su norma (longitud):

$$|\vec{v}| = \sqrt{\left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2} \qquad \text{(A.23)}$$

Así, el producto punto o producto interno entre dos vectores puede expresarse de dos modos equivalentes, uno utilizando componentes y otro utilizando el coseno del ángulo formado entre ellos:

$$\vec{v} \cdot \vec{w} = v^1 w^1 + v^2 w^2 + v^3 w^3 = |\vec{v}||\vec{w}|\cos\theta \qquad \text{(A.24)}$$

Una cosa muy bonita entre senos y cosenos es que son derivadas (e integrales) entre sí.

$$\frac{d}{d\theta}\text{sen}\theta = \cos\theta \qquad \text{(A.25)}$$

$$\frac{d}{d\theta}\cos\theta = -\text{sen}\theta \qquad \text{(A.26)}$$

Lo único que tiene que recordar es dónde va el signo menos, que puede deducir si recuerda que el cosθ es el que decrece a partir de uno, por lo que su derivada para ángulos pequeños y positivos tiene que ser negativa, lo cual se corresponde con –senθ. Las integrales siguen un patrón similar, desplazando el signo menos (cosa que tiene sentido si recordamos que las integrales anulan las derivadas).

$$\int \operatorname{sen}\theta \, d\theta = -\cos\theta \qquad \text{(A.27)}$$

$$\int \cos\theta \, d\theta = \operatorname{sen}\theta \qquad \text{(A.28)}$$

1 Dos matemáticos están en un bar, discutiendo sobre la competencia matemática del público en general. Uno de ellos se muestra extremadamente despectivo, mientras que el otro insiste en que los profanos pueden ser sorprendentemente entendidos. En un momento en que el primer matemático va al baño, el segundo aprovecha para llamar a la camarera. «Escuche», le dice, «cuando vuelva mi amigo, le haré a usted una pregunta, a la que usted responderá: "Un tercio de x al cubo", ¿de acuerdo? No se preocupe de lo que esto signifique, usted limítese a decir "un tercio de x al cubo". No lo olvide». La camarera asiente con la cabeza y va practicando las palabras mientras se retira: «Un tercio… de x … al cubo…».

Cuando el primero matemático regresa, el segundo vuelve a llamar a la camarera. «¿Le importaría ayudarnos un momento? Mi amigo no cree que la gente en general sepa gran cosa de matemáticas. Déjeme que le pregunte: ¿Cuál es la integral de x al cuadrado?»

«Un tercio de x al cubo», proclama la camarera orgullosa.

El primer matemático se queda impresionado, y tiene que reconocer que quizá finalmente su amigo tenga razón.

Mientras la camarera se retira, se gira hacia ellos con una sonrisa y añade: «Más una constante».

APÉNDICE B:

CONEXIONES Y CURVATURA

En el Capítulo 7, dedicado a la geometría, recorrimos todas las ideas necesarias para entender los conceptos referentes a las líneas geodésicas y a la ecuación de Einstein. Pero no incluimos absolutamente todos los pasos que se necesitarían para calcular tales cosas para una métrica dada. Rellenaremos aquí los espacios en blanco. Utilizaremos índices griegos, en lugar de latinos, imaginando que estamos en un espacio-tiempo de cuatro dimensiones, pero las fórmulas funcionan igual de bien en el espacio o en otras dimensiones de diferente número.

Mientras progresábamos hacia la ecuación de Einstein en el Capítulo 8, tuvimos que definir el escalar de curvatura de Ricci, en que intervenía la idea de «métrica inversa». Intentemos ser un poco más explícitos acerca de lo que esto significa. Comenzaremos por introducir un tensor extremadamente útil, la delta de Kronecker, que tiene un superíndice y un subíndice. En cuatro dimensiones, su aspecto es algo así:

$$\delta^{\mu}{}_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{cases} 1, & \mu = \nu \\ 0, & \mu \neq \nu \end{cases} \qquad \text{(B.1)}$$

Pensada como una matriz, la delta de Kronecker es simplemente la matriz identidad. La matriz identidad desempeña, en el terreno de las matrices, el

mismo papel que el 1 en aritmética. Cuando multiplicamos cualquier matriz por la matriz identidad, obtenemos la misma matriz.

Una vez entendido esto, podemos concebir la matriz inversa como el tensor por el que deberíamos multiplicar la métrica para obtener la delta de Kronecker. El tensor métrico gμν es un tensor simétrico con dos subíndices, de modo que la métrica inversa será un tensor simétrico con dos superíndices, gρσ, que cumple:

$$g^{\mu\lambda} g_{\lambda\nu} = \delta^{\mu}{}_{\nu} \qquad \text{(B.2)}$$

Deberíamos detenernos un momento a considerar lo que sucede aquí con los índices. Hay dos tipos de índices en la expresión de un tensor: índices libres e índices mudos. Los índices mudos son aquellos que aparecen dos veces, una vez arriba y la otra abajo, por lo que se suman, como sucede con λ en (B.2). No importa la letra que usemos para representarlos, con tal de que sea siempre la misma, y tiene que haber justamente un índice mudo arriba y un índice mudo abajo. (No podemos sumar índices repetidos que estén ambos arriba o ambos abajo). Los índices libres, por el contrario, aparecen solo una vez en cada expresión, como sucede con μ y ν en (B.2). Las letras que elijamos para representar los índices libres tampoco importan, pero es de importancia crucial que coincidan: cada término (es decir, cada producto de tensores) en cualquier ecuación debe tener los mismos índices libres. Vemos esto en (B.2), donde tanto en la parte izquierda como en la derecha aparecen una μ arriba y una ν abajo como índices libres. Si está intentando sumar expresiones de tensores con índices libres no coincidentes, es que hay algo que ha salido terriblemente mal.

La geometría euclidiana corriente utiliza implícitamente una métrica, solo que no se la han explicado. El producto punto, por ejemplo, entre dos vectores euclidianos tridimensionales es . Pero los elementos de esa métrica euclidiana plana (en coordenadas cartesianas) son:

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{(B.3)}$$

Comparando (las versiones tridimensionales de) (B.1) y (B.2), los elementos de la métrica inversa tienen precisamente el mismo aspecto:

$$g^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{(B.4)}$$

Esta es la razón por la que usted puede muy bien haber cursado niveles completos de enseñanza secundaria sobre geometría sin escuchar siquiera la palabra «métrica». Estaba ahí, pero siempre de modo implícito, nunca expresada explícitamente, porque usted seguía anclado en el espacio plano de coordenadas cartesianas, y los elementos de la métrica, de la métrica inversa y de la delta de Kronecker eran los mismos.

Esto no puede generalizarse, los elementos de la métrica inversa no suelen ser los mismos que los de la métrica. Si la métrica es diagonal, nos encontramos ante la feliz situación de que los elementos de la métrica inversa sin simplemente los recíprocos de los elementos de la métrica. (Si la métrica no es diagonal, las cosas se complican rápidamente). Por ejemplo, en un espacio plano tridimensional euclidiano, con coordenadas esféricas, la métrica toma la forma:

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2(\text{sen}\theta)^2 \end{pmatrix} \qquad \text{(B.5)}$$

Esto implica que la métrica inversa tiene los elementos:

$$g^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^{-2} & 0 \\ 0 & 0 & r^{-2}(\text{sen}\theta)^{-2} \end{pmatrix} \qquad \text{(B.6)}$$

En un espacio plano, tenemos al menos la opción de utilizar coordenadas cartesianas en que la métrica y la métrica inversa son esencialmente las mismas, pero en situaciones más generales no existe tal opción, por lo que es importante mantener la distinción de conceptos.

La presencia de la métrica y la métrica inversa permite un bonito método de manipulación de tensores: subir índices y bajar índices. Que una posición de índice sea subíndice o superíndice importa, como muestra el ejemplo de la propia métrica. Pero podemos convertir un índice superior en inferior multiplicando por la métrica y compensando ese índice, y del mismo modo podemos convertir un índice inferior en superior con la métrica inversa. Dado un vector vμ, por ejemplo, podemos bajar su índice mediante:

$$v_\mu = g_{\mu\nu} v^\nu \qquad \text{(B.7)}$$

El hecho de que la métrica inversa cumpla (B.2) nos garantiza que podemos bajar un índice y luego subirlo de nuevo, volviendo así al mismo tensor con

que habíamos comenzado (porque compensar δμν es lo mismo que no hacer nada):

$$g^{\mu\lambda}v_{\lambda} = g^{\mu\lambda}g_{\lambda\nu}v^{\nu} = \delta^{\mu}{}_{\nu}v^{\nu} = v^{\mu} \qquad \text{(B.8)}$$

Este mínimo de tecnología de tensores era necesario para definir el escalar de curvatura del Ricci en el Capítulo 8. El tensor de Riemann viene equipado de fábrica con un único índice superior y tres índices inferiores, por lo que resulta sencillo «reducir» (compensar) un índice para definir el tensor de Ricci: Rμν = Rλμλν. Pero entonces nos quedamos atorados en un tensor con dos índices inferiores: no podemos realizar otra reducción para obtener un escalar. Lo que podemos hacer es subir un índice utilizando la métrica inversa, Rμν = gμλRλν. Ahora, esto puede reducirse para definir el escalar de curvatura: R = Rλλ, lo que equivale a R = gλσRλσ. Entonces, si es usted Einstein, puedes poner esto en funcionamiento para definir un tensor que pueda ser proporcional al tensor energía-momento y seguir preservando el principio de conservación de la energía.

También hay un aumento de índice secreto que fue crucial, para empezar, en el proceso de definición del tenso de Riemann. Desempeñó un papel central el transporte en paralelo de un vector Wμ a lo largo de una trayectoria parametrizada xμ (λ). (A pesar de ser una letra griega, λ no es aquí un índice, sino el parámetro que nos dice en qué lugar del camino estamos). Esto significa que tenemos que definir un valor del vector para cada punto, Wμ (λ), que cumpla la ecuación del transporte paralelo.

$$\frac{d}{d\lambda}W^{\mu} + \Gamma^{\mu}{}_{\sigma\rho}\frac{dx^{\sigma}}{d\lambda}W^{\rho} = 0 \qquad \text{(B.9)}$$

Entendemos el sentido de la mayor parte de esta notación, pero necesitamos definir Γμσρ. Se conocen como coeficientes de conexión o símbolos de Christoffel. Parecen sin duda elementos de un tensor, pero, estrictamente hablando, no lo son, por eso los llamamos «coeficientes» o simplemente

«símbolos». (La razón es que dependen de coordenadas en un sentido no tensorial). Estos coeficientes definen aquello que entendemos en la práctica por la conexión en una variedad, la información que necesitamos para comparar vectores y tensores ubicados en puntos cercanos. La idea de conexión desempeña también un papel importante en teorías de gauge, en física de partículas.

Para definir los coeficientes de conexión, introduciremos más nomenclatura, pero esta vez es simplemente una estrategia para el ahorro de trabajo, más que un giro conceptual nuevo. En el estudio de campos tensoriales en variedades, es muy común tomar derivadas parciales con respecto a las coordenadas xμ. Tan común que hemos inventado una notación precisa ad hoc:

$$\frac{\partial}{\partial x^{\mu}} = \partial_{\mu} \qquad \text{(B.10)}$$

Ya ve la añagaza: xμ tiene un índice superior, pero en la derivada parcial aparece en el denominador, por lo que operador de la derivada parcial $\partial\mu$ aparece con un índice inferior.

Ahora que entendemos tanto la métrica inversa como la notación de la derivada parcial, podemos presentar la fórmula para los coeficientes de conexión:

$$\Gamma^{\rho}{}_{\mu\nu} = \frac{1}{2} g^{\rho\lambda} \left(\partial_{\mu} g_{\nu\lambda} + \partial_{\nu} g_{\lambda\mu} - \partial_{\lambda} g_{\mu\nu} \right) \qquad \text{(B.11)}$$

Es muy posible que mientras usted está leyendo esto, en algún lugar del mundo haya estudiantes ocupados en la relatividad general que estén calculando coeficientes de conexión para alguna métrica determinada y que estén utilizando esta ecuación. Le animamos a intentarlo. Si utiliza la métrica plana con coordenadas esféricas (B.5), es lo bastante difícil como

para resultar interesante, sin sobrepasar el límite de lo abordable. Puesto que hay tres índices en Γρμν, con tres dimensiones, corresponden $3^3 = 27$ elementos. Pero con una métrica diagonal que solo depende de dos de las coordenadas, muchos de esos elementos terminarán siendo cero. Tenga presentes, eso sí, todos los índices que se compensan.

Los coeficientes de conexión definen el funcionamiento del transporte paralelo, y por tanto definen también líneas geodésicas, que a fin de cuentas no son sino trayectorias que transportan en paralelo su propios vectores de velocidad dxμ/dλ. Sustituyendo esto por Wμ en (B.9), obtenemos la ecuación geodésica:

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu{}_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0 \qquad \text{(B.12)}$$

Dada una métrica gμν, podemos calcular los coeficientes de conexión a partir de (B.11), y luego utilizar esta ecuación para resolver geodésicas xμ (λ). Esto nos mostrará el modo en que los objetos físicos reales, desde planetas hasta fotones, viajan libremente a través de un espacio-tiempo así.

Finalmente, el otro uso importante de los coeficientes de conexión es el de definir el tensor de curvatura de Riemann. Conceptualmente lo explicamos en el Capítulo 7, pero en algún momento tendrá que sentarse a calcular los elementos. Aquí tiene la fórmula en todo su esplendor:

$$R^\rho{}_{\sigma\mu\nu} = \partial_\mu\Gamma^\rho{}_{\nu\sigma} - \partial_\nu\Gamma^\rho{}_{\mu\sigma} + \Gamma^\rho{}_{\mu\lambda}\Gamma^\lambda{}_{\nu\sigma} - \Gamma^\rho{}_{\nu\lambda}\Gamma^\lambda{}_{\mu\sigma} \qquad \text{(B.13)}$$

La verdad es que, hoy en día, los alumnos no suelen calcular a mano los elementos del tensor de Riemann. Hay programas de ordenador que harán el trabajo por usted. Nunca vivirán la experiencia formativa de quedarse hasta altas horas de la noche en compañía de las hojas de borrador desparramadas por la mesa de una cocina, llenas de símbolos griegos,

intentando detectar el paso en que pusiste por error una μ en lugar de una ν.
Tiempos felices.